



VYSOKÁ ŠKOLA BÁŇSKÁ – TECHNICKÁ UNIVERZITA OSTRAVA  
EKONOMICKÁ FAKULTA

KATEDRA FINANČÍ

Modelování pojistných škod v rámci havarijního pojištění  
Modelling of Claim Severity for a Motor Hull Insurance Portfolio

Student:	Bc. Šárka Marčíková
Vedoucí diplomové práce:	Ing. Jiří Valecký, Ph.D.

Ostrava 2016

## Zadání diplomové práce

Student: **Bc. Šárka Marčíková**

Studijní program: **N6202 Hospodářská politika a správa**

Studijní obor: **6202T010 Finance**

Téma: **Modelování pojistných škod v rámci havarijního pojištění  
Modelling of Claim Severity for a Motor Hull Insurance Portfolio**

Jazyk vypracování: **čeština**

Zásady pro vypracování:

1. Úvod
  2. Řízení pojistných rizik
  3. Charakteristika použité metodiky
  4. Odhad a vyhodnocení regresního modelu
  5. Závěr
- Seznam použité literatury  
Seznam zkratk  
Prohlášení o využití výsledků diplomové práce  
Seznam příloh  
Přílohy

Seznam doporučené odborné literatury:

HARDIN, James W. and Joseph M. HILBE. *Generalized Linear Models and Extensions*. 2nd ed. College Station: Stata Press, 2007. 387 s. ISBN 978-1-59718-014-6.

JONG, Piet de and Gillian Z. HELLER. *Generalized Linear Models for Insurance Data*. Cambridge: Cambridge University Press, 2008. 196 s. ISBN 978-0-521-87914-9.

OHLSSON, Esbjörn and Björn JOHANSSON. *Non-Life Insurance Pricing with Generalized Linear Models*. Berlin: Springer, 2010. 174 s. ISBN 978-3-642-10790-0.

Formální náležitosti a rozsah diplomové práce stanoví pokyny pro vypracování zveřejněné na webových stránkách fakulty.

Vedoucí diplomové práce: **Ing. Jiří Valecký, Ph.D.**

Datum zadání: 20.11.2015

Datum odevzdání: 22.04.2016



  
Ing. Iveta Ratmanová, Ph.D.  
vedoucí katedry

  
prof. Dr. Ing. Dana Dluhošová  
děkanka fakulty

Prohlašuji, že jsem celou práci, včetně všech příloh, vypracovala samostatně.

V Ostravě dne 22. dubna 2016



.....

Bc. Šárka Marčíková

Ráda bych poděkovala Ing. Jiřímu Valeckému, Ph.D. za vstřícný přístup, odbornou pomoc a cenné rady, kterými přispěl k vypracování této diplomové práce.

Tato diplomová práce vznikla za podpory projektu SGS SP2015/75 Aplikace zobecněných lineárních modelů v pojišťovnictví a ve financích.

## **OBSAH**

<b>1</b>	<b>Úvod.....</b>	<b>5</b>
<b>2</b>	<b>Řízení pojistných rizik .....</b>	<b>7</b>
2.1	Vymezení rizika v rámci pojistné teorie.....	7
2.1.1	Pojem riziko .....	7
2.1.2	Klasifikace rizik .....	8
2.1.3	Pojistná rizika.....	9
2.2	Pojištění a riziko .....	11
2.2.1	Základní klasifikace pojištění.....	11
2.2.2	Havarijní pojištění .....	12
2.3	Risk management .....	13
2.3.1	Identifikace rizika.....	14
2.3.2	Ohodnocení a kvantifikace rizika.....	15
2.3.3	Kontrola a řízení rizika.....	16
2.4	Regulační pojetí risk managementu v pojišťovnictví.....	17
2.4.1	Legislativní rámec risk managementu.....	18
2.4.2	Koncept Solvency II.....	19
2.5	Aktuální stav risk managementu v pojišťovnictví.....	21
<b>3</b>	<b>Charakteristika použité metodiky .....</b>	<b>23</b>
3.1	Zobecněné lineární modely .....	23
3.2	Metoda maximální věrohodnosti .....	25
3.3	Jednorozměrná analýza pomocí Z-testu .....	26
3.4	Regresní gamma model .....	27
3.4.1	Gamma rozdělení pravděpodobnosti.....	28
3.4.2	Odvození gamma modelu.....	29
3.4.3	Odhad parametrů gamma modelu .....	30
3.5	Teorie extrémních hodnot.....	31
3.5.1	Zobecněné rozdělení extrémních hodnot .....	33

3.5.2	Přístup Peak over the threshold .....	34
3.6	Regresní zobecněný Paretův model .....	35
3.6.1	Zobecněné Paretovo rozdělení .....	35
3.6.2	Odvození zobecněného Paretova modelu .....	36
3.6.3	Odhad parametrů zobecněného Paretova modelu .....	36
3.7	Regresní smíšený model .....	38
3.8	Reziduální analýza .....	39
<b>4</b>	<b>Odhad a vyhodnocení regresního modelu .....</b>	<b>41</b>
4.1	Analýza vstupních dat .....	41
4.2	Konstrukce regresního log-gamma modelu .....	43
4.2.1	Jednorozměrná analýza na bázi Z-testu .....	43
4.2.2	Odhad parametrů modelu .....	45
4.3	Konstrukce regresního smíšeného modelu .....	46
4.3.1	Odhad parametrů log-gamma modelu .....	47
4.3.2	Odhad parametrů zobecněného Paretova modelu .....	48
4.4	Porovnání regresních modelů .....	49
4.4.1	Analýza Pearsonových reziduí .....	50
4.4.2	Analýza hrubých reziduí .....	53
4.4.3	Analýza směrodatné odchylky .....	57
4.5	Souhrnné zhodnocení regresních modelů .....	57
<b>5</b>	<b>Závěr .....</b>	<b>59</b>
	<b>Seznam použité literatury .....</b>	<b>62</b>
	<b>Seznam zkratk .....</b>	<b>64</b>
	<b>Prohlášení o využití výsledků diplomové práce</b>	
	<b>Seznam příloh</b>	
	<b>Přílohy</b>	



# 1 Úvod

Proces finančního modelování je v posledních letech nedílnou součástí řízení finančních institucí či jiných velkých korporací, neboť výrazným způsobem napomáhá při rozhodování osob v řídicích orgánech jednotlivých subjektů. Díky neustálému rozvoji informačních technologií a systémů se na finanční modely nahlíží jako na nástroj umožňující vyřešit složité finanční či ekonomické otázky.

Vzhledem ke skutečnosti, že v odvětví pojišťovnictví se dennodenně setkáváme s prvkem náhody a jejími důsledky, je vhodné, aby pojistitelé byli především náležitě připraveni na různé scénáře vyplývající z výkonu pojišťovací činnosti a aby byli schopni zabezpečit ochranu svých klientů. Důležitou oblastí sledovanou pojišťovnami je pravděpodobnost výskytu pojistné události a výše škody s ní související, jelikož tyto aspekty mohou mít bezprostřední vliv na solventnost pojišťovny a její celkovou pozici na trhu. Snahou každého pojistitele by pak mělo být zohlednění veškerých podstupovaných rizik a jejich případných důsledků tak, aby nebyla žádným způsobem ohrožena existence samotné instituce, ale ani závazky vůči jejím klientům.

Cílem této diplomové práce je vytvořit a odhadnout regresní model pro modelování individuálních pojistných škod (tj. pro každou smlouvu) daného pojistného kmene. Model musí být dostatečně přesný a přesnost musí být vyhodnocena.

Diplomová práce je koncipována do pěti kapitol, kdy první kapitolou je úvod a poslední kapitolou závěr. Teoretické části práce je věnována druhá a třetí kapitola, praktická část práce je obsažena v kapitole čtvrté.

Druhá kapitola je zaměřena na problematiku řízení pojistných rizik. Nejdříve se zabýváme vymezením pojistného rizika a jeho návaznosti na pojištění, blíže se pak soustředíme na popis risk managementu pojišťoven. Dále je zde stručně charakterizován regulační rámec řízení rizik a také aktuální situace risk managementu v odvětví pojišťovnictví.

Třetí kapitola pojednává o teoretických základech použité metodologie. V prvních odstavcích této části jsou přiblíženy principy fungování zobecněných lineárních modelů, poté jsou zde samostatně nastíněny metoda maximální věrohodnosti a jednorozměrná analýza. V návaznosti na popsané metody je dále charakterizován regresní gamma model. V další části se pak věnujeme teorii extrémních hodnot a konstrukci regresního zobecněného Paretova modelu dle přístupu Peak over the threshold. Závěrem kapitoly je stručně charakterizována jedna z metod statistické verifikace modelu, konkrétně analýza reziduí.

Ve čtvrté kapitole je obsažen vlastní postup řešení vedoucí k vytvoření a odhadu regresního modelu zachycujícího vztah mezi velikostí dané pojistné škody a individuálními rizikovými faktory. V úvodní části této kapitoly jsou vymezena vstupní data společně s jejich základními statistickými charakteristikami. Dále se zde věnujeme odhadu parametrů zobecněného lineárního modelu za předpokladu gamma rozdělení. Druhým uvažovaným modelem, jehož parametry jsou odhadovány, je tzv. smíšený model, který je konstruován za předpokladu smíšeného rozdělení pravděpodobnosti skládajícího se z gamma rozdělení a zobecněného Paretova rozdělení. Po odhadu parametrů obou regresních modelů metodou maximální věrohodnosti analyzujeme přesnost těchto odhadů prostřednictvím grafického zachycení reziduí. Na závěr jsou oba modely zhodnoceny a je zvolen ten model, který lépe odpovídá pozorovaným datům.

## **2 ŘÍZENÍ POJISTNÝCH RIZIK**

Snahou pojišťoven, jakožto podnikatelských subjektů, je generovat zisk a obstát v silné konkurenci na pojistném trhu. Avšak pro úspěšné řízení pojišťovny je nezbytné brát v potaz i rizika, kterým tato finanční instituce čelí. Pojišťovny proto neustále usilují o zkvalitnění procesu řízení rizik, což by mělo mít pozitivní vliv na jejich prosperitu a rozvoj.

V této kapitole jsou nejdříve stručně charakterizována rizika v oblasti pojišťovnictví, poté je zde popsán vztah mezi rizikem a pojištěním, přičemž se detailněji zaměřujeme pouze na havarijní pojištění, další odstavce jsou věnovány samotnému procesu řízení rizik, tedy risk managementu, a vymezení jeho regulačního rámce v pojišťovnictví, závěrem druhé kapitoly je popsán současný stav risk managementu v této oblasti finančního trhu. Hlavním zdrojem informací jsou publikace Cipra (2002), Ducháčková (2009), Ducháčková, Daňhel (2010), Řezáč (2011).

### **2.1 VYMEZENÍ RIZIKA V RÁMCI POJISTNÉ TEORIE**

Téměř při jakékoliv každodenní činnosti, kterou vykonáváme, se můžeme setkat s rozličnými riziky ohrožujícími výsledek naší práce. Nejinak je tomu na pojistném trhu, jež je velmi úzce spojován s nahodilostí a důsledky nahodilosti. Nahodilé jevy mohou mít pro dané subjekty a celkově pro lidskou společnost kladné, ale také negativní důsledky, a proto je nutná neustálá snaha správně předvídat a eliminovat možná nebezpečí včetně jejich negativních důsledků.

#### **2.1.1 POJEM RIZIKO**

V odborné literatuře je k dispozici celá řada definic týkajících se rizika. Riziko je provázáno s nejednoznačností výsledků jevů a procesů, obecně jej lze chápat jako nejistotu spojenou se vznikem určité potenciální situace.

Dle Ducháčková (2009) lze tvrdit, že z pohledu pojistné teorie je riziko nejistota, která se dá kvantifikovat, a to pomocí předem známého objektivního rozdělení pravděpodobnosti náhodných veličin počtem pravděpodobnosti. Konkrétně můžeme riziko charakterizovat jako možnost vzniku události s výsledkem odchylicím se od cíle s určitou objektivní pravděpodobností.

Jak uvádí Ducháčková, Daňhel (2010), v rámci pojišťovnictví je možné riziko interpretovat různými způsoby. Jednak si pod pojmem riziko můžeme představit předmět ohrožovaný nahodilým nebezpečím, tedy předmět pojištění. Dále lze, z matematického pohledu, nahlížet na riziko jako na pravděpodobnost vzniku nahodilé události s negativními

dopady na daný ekonomický subjekt. Třetím možným výkladem rizika je náhodná událost zapříčiňující škodu. Takto specifikovaná rizika jsou pak na pojistném trhu přesouvána prostřednictvím pojištění na speciální instituci – pojišťovnu.

### **2.1.2 KLASIFIKACE RIZIK**

Ke způsobům členění rizik přistupují autoři odborných publikací poměrně rozdílně, navíc existuje celá škála kritérií klasifikace rizik. Pro účely této diplomové práce jsou níže shrnuty pouze základní kategorie rizik používané v oblasti pojišťovnictví.

*Z pohledu možných důsledků* rozlišujeme riziko čisté a riziko spekulativní. O čistém riziku hovoříme v případě, že dochází ke vzniku pouze negativní odchylky od předpokládaného cíle, lidé jej tedy záměrně nepodstupují. Výhradním předmětem pojištění jsou právě čistá rizika, jejichž realizací vznikají náhodné potřeby. Naopak spekulativní neboli záměrné riziko je charakteristické možností záporné i kladné odchylky od cíle, přičemž daný subjekt se dobrovolně vystavuje riziku, a to např. při hraní hazardních her, sázení či při spekulacích na burze.

*Podle příčiny vzniku* můžeme rizika rozdělit na přírodní rizika a rizika vyvolaná lidským faktorem. Přírodní rizika jsou spjatá s působením přírodních sil, přičemž jejich vznik a vývoj je velmi těžké předvídat a ovlivnit. Rizika způsobená lidskou činností jsou dále členěna na rizika vyvolaná lidmi a rizika technická, jejichž příčinou jsou technické prostředky stvořené lidmi.

*Z hlediska možnosti sjednat si na dané riziko pojištění* členíme rizika na pojistitelná a nepojistitelná. S neustálým a rychlým rozvojem informačních technologií a moderních technik dochází postupně k posouvání hranice pojistitelnosti rizika, která je nicméně výrazně závislá na peněžních prostředcích. Obecně můžeme tvrdit, že pojistit lze téměř vše, pokud máme k dispozici dostatek financí, avšak na druhou stranu daná pojišťovna má právo se rozhodnout, zdali pojistné riziko přijme. Dle Řezáč (2011) musí pojistitelná rizika splňovat požadavky identifikovatelnosti rizika (jednoznačné určení příčiny pojistné události), vyčíslitelnosti škody, nahodilosti vzniku pojistné události a ekonomické přijatelnosti rizika pro pojišťovnu.

*Na základě metodiky Solvency II* můžeme rizika v oblasti pojišťovnictví rozdělit na pojistně technické riziko, tržní riziko, úvěrové riziko, riziko likvidity a operační riziko, přičemž struktura těchto rizik je u životních a neživotních pojišťoven odlišná. Pro životní pojišťovny je velmi důležitá orientace především na rizika tržní a úvěrová, zatímco neživotní pojišťovny by měly svou pozornost věnovat pojistně technickým rizikům.

Dle sféry výskytu klasifikujeme rizika na finanční rizika a pojistná rizika. Pojistná rizika jsou typická pro trh pojištění a jsou blíže specifikována v následující kapitole.

Výše uvedený výčet členění rizik není zdaleka úplný, kategorií rizik existuje skutečně velmi mnoho a každý ekonomický subjekt čelí vybraným rizikům podle zaměření a způsobu podnikání. Komerční pojišťovny jsou samozřejmě vystavovány jednak tzv. obchodně-podnikatelským rizikům, kam spadá např. kreditní či tržní riziko, ale také rizikům charakteristickým pro toto odvětví, a to pojistným rizikům.

### **2.1.3 POJISTNÁ RIZIKA**

Pojistné riziko můžeme definovat jako míru pravděpodobnosti vzniku pojistné události vyvolané pojistným nebezpečím. Pojistné riziko je rizikem čistým a je možné jej přenést na pojistitele prostřednictvím pojištění. Je nezbytné, aby pojistné riziko bylo vždy jasně ohraničeno rámcem pojistného práva, pojistnou dobou a místním vymezením. Každou pojišťovnou jsou pak vypracovány příslušné pojistné podmínky dle druhu pojištění. Pojistná rizika lze dále rozdělit na pojistně technická rizika a specifická rizika.

#### **A) POJISTNĚ TECHNICKÁ RIZIKA**

*„Pojistně technické riziko pojišťovny je definováno jako možnost vzniku kladné či záporné odchylky od pojišťovnou apriori vypočtených úplných vlastních nákladů pojišťovny, tj. od pojišťovnou předpokládaného škodního průběhu a nákladů správní režie“, viz Ducháčková, Daňhel (2010, str. 49).* Pojistně technické riziko lze tedy jinými slovy formulovat jako potencionální nebezpečí, že u určité pojišťovny ve skutečnosti nedojde k vyrovnání mezi přijatým pojistným a vyplaceným pojistným plněním. Tento typ rizika je spjat s hospodařením pojišťovny a jeho existence je důsledkem faktu, že pojišťovny provozují služby nahodilého charakteru.

V rámci pojistně technického rizika dle Ducháčková (2009) rozlišujeme z hlediska původu odchylek od kalkulace:

- náhodné riziko,
- riziko změn,
- riziko omylu,
- riziko načasování.

*Náhodným rizikem* se rozumí náhodné kolísání kolem očekávaného škodního průměru, který se v čase nemění. Náhodné riziko lze dále rozdělit dle jeho rozsahu na normální náhodné riziko, jehož škodní průběh vykazuje běžné odchylky od očekávaného průměru a je

snesitelné pro hospodaření pojišťovny, a katastrofální náhodné riziko, u kterého je škodní průběh ovlivňován škodami značného rozsahu.

*Riziko změn* je zapříčiněno změnami výchozích podmínek, dle kterých bylo stanoveno pojistné. V případě této situace dochází ke změnám v rizikovosti, přičemž tyto změny mohou mít charakter cyklický (spojitost s průběhem hospodářského cyklu), trendový (návaznost na strukturální, biologické či klimatické změny) nebo nepravidelný (např. souvislost s výkyvy počasí).

*Riziko omylu* spočívá ve špatném odhadu škodního průběhu, což může mít za následek nesprávné určení ceny daného pojistného produktu pojišťovnou.

*Riziko načasování* je významné zejména u pojištění, která jsou specifická dlouhou dobou mezi vznikem pojistné události a výplatou pojistného plnění. Toto riziko je realizováno v okamžiku, kdy dojde k pojistné události a následně potřebě výplaty pojistného plnění podstatně dříve, než se očekávalo. Pojistitel obvykle ještě nemá vytvořeny dostatečné rezervy, a tudíž je pro něj dané pojištění ztrátové.

Jednotlivé pojišťovny si kladou za cíl eliminovat pojistně technické riziko na minimum. Tohoto cíle lze dosáhnout různými metodami a postupy. Jednou ze základních možností je vyrovnat pojistně technická rizika tzv. rozložením, v rámci kterého se daný pojistitel snaží pokrýt rizika na co nejrozsáhlejší území. Dalším možným řešením je diverzifikace rizik, tedy krytí co nejširší struktury rizik. Pojišťovnami často využívaným nástrojem snížení rizika je dále tvorba rezerv na krytí výkyvů škodního průběhu z hlediska věcného, časového a místního. Hlavním smyslem vytváření rezerv je vyrovnání časových nesouladů mezi příjmem pojistného a opožděnými výplatami pojistných plnění, případně krytí náhodných výkyvů ve škodním průběhu. V neposlední řadě mohou také pojistitelé redukovat pojistně technické riziko pomocí zajištění. Dle Cipra (2002) lze zajištění definovat jako přenesení části rizika, jež převzal pojistitel od pojištěných, na jiného organizátora pojištění – zajištěle. Zajišťovací činnost je téměř analogická činnosti pojišťovací až na to, že se odehrává na úrovni vztahu mezi pojišťovnou a zajišťovnou.

## **B) SPECIFICKÁ RIZIKA**

Specifická rizika lze chápat jako rizika, která jsou typická pro odvětví pojišťovnictví a která nespádají do oblasti pojistně technických rizik. Jako významná rizika této kategorie můžeme zmínit riziko investičního portfolia pojišťovny související s volbou vhodné investiční strategie, riziko zajištění spojené s možností nesolventnosti zajištěle či riziko solventnosti

vyplývající z neschopnosti pojistitele dostát svým závazkům z pojišťovací činnosti v dlouhodobém měřítku.

## **2.2 POJIŠTĚNÍ A RIZIKO**

Každý ekonomický subjekt si může zvolit jednu ze dvou možností, jak se vyrovnat s vzniklými nahodilými událostmi z hlediska finančního zajištění. Daná osoba může následky krýt z vlastních zdrojů nebo může využít pojištění. Z tohoto pohledu lze tedy pojištění označit jako nástroj finanční eliminace negativních důsledků nahodilosti, jak uvádí Ducháčková (2009). Nicméně důležité je si uvědomit, že pomocí pojištění nelze ovlivnit výskyt náhodných událostí, ale pouze jejich finanční dopad.

Pod pojmem pojištění si lze představit finanční službu, jejímž předmětem je poskytnutí pojistné ochrany za úplatu. Z právního hlediska lze pojištění definovat jako právní vztah, ve kterém na sebe pojistitel přebírá riziko a současně závazek poskytnout pojištěnému pojistné plnění, dojde-li k nahodilé události předem vymezené v příslušné pojistné smlouvě a v pojistných podmínkách, a pojistník se zároveň zavazuje za přenesení rizika platit danému pojistiteli pojistné dle smluvených podmínek.

### **2.2.1 ZÁKLADNÍ KLASIFIKACE POJIŠTĚNÍ**

I v případě pojištění existuje řada hledisek a kategorií, podle kterých lze pojištění členit. V odstavcích níže jsou vyjmenovány a stručně charakterizovány pouze ty hlavní a základní kategorie pojištění.

*Podle právního hlediska* se pojištění dělí na pojištění dobrovolné a pojištění povinné. Dobrovolné pojištění je sjednáváno z vlastní vůle zájemce o pojištění, kdežto uzavření povinného pojištění vyplývá z právních předpisů.

*Dle způsobu financování* se rozlišují pojištění sociální a pojištění komerční. Komerční pojištění je využíváno ekonomickými subjekty dobrovolně v návaznosti na jejich rozhodnutí a potřeby. Pro sociální pojištění je charakteristická povinnost pojištění, jelikož se jedná o zákonné pojištění a nesjednává se pojistnou smlouvou.

*Podle pojistného odvětví* se vymezují životní a neživotní pojištění, přičemž životní pojištění slouží pro krytí životních rizik (riziko smrti a dožití) a neživotní pojištění pro krytí různých druhů neživotních rizik (např. riziko odcizení majetku, riziko úrazu)

*Dle způsobu tvorby rezervy* jsou pojištění klasifikována na pojištění riziková a rezervotvorná. Pro rizikové pojištění platí podmíněná návratnost finančních prostředků určená vznikem pojistné události, nicméně u tohoto typu pojištění není jasné, zdali k pojistné události vůbec dojde během trvání pojištění, pojistné je proto spotřebováváno pojistitelem průběžně na



pokrytí výplaty ostatních pojistných plnění. Naopak v případě rezervotvorného pojištění víme s jistotou, že pojistná událost nastane, pouze nevíme v jakém časovém okamžiku, a tudíž je v rámci tohoto pojištění vytvářena rezerva na výplatu sjednaných pojistných plnění v budoucnosti.

*Podle způsobu krytí potřeb* se pojištění člení na obnosová a škodová. Obnosová pojištění neboli pojištění na pojistnou částku jsou charakteristická stanovením konkrétní pojistné částky udávající absolutní výši pojistného plnění, přičemž reálná výše škody se nezjišťuje, jelikož ji nelze ani přesně peněžně vyjádřit. U škodových pojištění je rozsah pojistného plnění navázán na velikost skutečně vzniklé škody, jejíž rozsah lze poměrně dobře odhadnout a vyčíslit.

### **2.2.2 HAVARIJNÍ POJIŠTĚNÍ**

Jelikož obsahem praktické části této diplomové práce je modelování pojistných škod v rámci havarijního pojištění, je vhodné, abychom zde přiblížili hlavní znaky a princip fungování havarijního pojištění.

Havarijní pojištění řadíme mezi neživotní pojištění, konkrétně spadá do kategorie pojištění majetku obyvatelstva. Toto pojištění lze dále specifikovat dle výše zmíněného členění jako pojištění komerční, dobrovolné, rizikové a škodové. Prostřednictvím havarijního pojištění jsou kryta tzv. havarijní rizika, konkrétně se jedná o rizika vzniku majetkových škod na dopravním prostředku, ale také na zboží přepravovaném dopravním prostředkem. Základní variantou pojistného krytí v oblasti havarijního pojištění je pokrytí pojistného nebezpečí havárie (nárazu). Pojišťovny ale také velmi často v rámci tohoto pojištění nabízejí možnost krytí dalších pojistných nebezpečí, kupříkladu nebezpečí živelní, kolize, odcizení, vandalství atd. V souvislosti s havarijním pojištěním je dále velmi často uplatňováno doplňkové připojištění asistenčních služeb.

Konstrukce havarijního pojištění se opírá o princip pojištění na novou hodnotu, což znamená, že pojištěnému se vyplácí pojistné plnění ve velikosti odpovídající nákladům na opravu, pokud tyto náklady nepřesahují cenu vozidla před pojistnou událostí. Dalším typickým znakem konstrukce havarijního pojištění je uplatňování excedentní franšízy (odečtení určité částky od pojistného plnění) nebo procentní spoluúčasti (účast pojištěného na úhradě škody prostřednictvím stanoveného procenta), neboť se tím snižuje výše pojistného plnění a ovlivňuje velikost pojistného. Ve velikosti pojistného v havarijním pojištění se odráží vybrané faktory sledované pojistiteli. Mezi tato kritéria patří hlavně typ a značka vozidla, pořizovací cena vozidla, stáří vozidla, zvolená pojistná nebezpečí, zvolená spoluúčast, způsob



platby, původ výroby vozidla, rizikovost řidičů užívajících vozidlo, bonus či malus a další, blíže viz Ducháčková (2009).

## **2.3 RISK MANAGEMENT**

Risk management lze považovat za dynamicky se rozvíjející vědní disciplínu, která nemá příliš dlouhou tradici, nicméně v poslední době je jí věnováno stále více pozornosti, a to nejen na manažerské úrovni pojišťoven, ale i dalších podnikatelských subjektů. Risk management neboli řízení rizik je v současné době jednou z klíčových oblastí řízení pojišťoven, neboť tyto instituce na sebe dennodenně přebírají velké množství rozličných rizik, které na ně převádí jejich klientela, a je proto nezbytné s nimi umět pracovat a ovládat je správným směrem.

*„Předmětem risk managementu je umožnit v tržní ekonomice zahrnutí projevů rizika, vyplývajících z nejednoznačnosti průběhu reálných ekonomických procesů, do rozhodování o hospodářských záležitostech“*, jak uvádí Ducháčková (2009, str. 21). Tato disciplína je tedy hlavně zaměřena na soustavnou analýzu ekonomické činnosti z hlediska zřetelných, potenciálních a také skrytých rizik. Hlavním cílem risk managementu je dosažení bezpečného fungování daného subjektu při co nejnižších nákladech vynaložených na zajištění této bezpečnosti. Rizikovní manažeři se také snaží zajistit dostatečné rezervy pro případ vzniku krizových situací v takové výši, aby nebyla výrazným způsobem ovlivněna celková stabilita podniku. Opatření přijatá oddělením řízení rizik by samozřejmě měla být v souladu s hlavními cíli, celkovou strategií a koncepcí dané pojišťovny. Úspěšnost oddělení risk managementu je závislá na kvalitách a zkušenostech jednotlivých manažerů a jejich postoji k riziku.

Risk management lze charakterizovat jako komplexní proces řízení zahrnující celou řadu činností, jež by měly směřovat k minimalizaci ztrát vyplývajících z možné realizace rizika. Rizikový manažer potřebuje k vykonávání této pozice znalosti z různých odvětví, jako např. z oblasti managementu, financí, oceňování, pojišťovnictví, statistiky, marketingu, právních norem atd. Jeho hlavní úkoly spočívají v maximálně možném omezení pravděpodobnosti výskytu rizik a zároveň v předvídání a organizování případných důsledků realizace rizik tak, aby jejich dopad byl pro podnik co nejnižší. Jednání risk managementu by mělo vést k nalezení a posouzení nejvýznamnějších rizik a následně k rozhodnutí, jak se s těmito riziky vypořádat, viz Ducháčková (2009).

Proces řízení rizika je možné rozdělit do tří základních fází – identifikace rizika, ohodnocení a kvantifikace rizika, kontrola a řízení rizika, přičemž tyto fáze se neustále

opakují a jsou vzájemně úzce propojeny. Jednotlivé činnosti jsou přiblíženy v následujících kapitolách.

### **2.3.1 IDENTIFIKACE RIZIKA**

První oblastí, kterou se zabývají rizikovní manažeři a kterou lze označit jako klíčovou činnost risk managementu, je identifikace rizika. Tato činnost spočívá v analýze rizikového prostředí a následném určení rizik, jež mohou ohrožovat ekonomickou stabilitu daného subjektu a jeho úspěšnost. Velmi důležitým aspektem je správná a hlavně včasná specifikace rizik a rizikových faktorů, jelikož subjekt je mnohem lépe schopen čelit těm rizikům, která dopředu zná a může se na jejich případnou realizaci určitým způsobem připravit. V rámci této první fáze risk managementu je třeba počítat s riziky zjevnými, hmatatelnými, ale i se skrytými, jejichž existence se zdá být vyloučena. Rizikovní manažeři by rovněž neměli opomínat neustálé sledování vývoje podnikatelského okolí, neboť v důsledku měnícího se ekonomického, sociálního i legislativního prostředí může dojít ke vzniku nových rizik.

Dle Řezáč (2011) si pod pojmem identifikace rizika můžeme představit zmapování celkové situace daného subjektu, přičemž se soustředíme na újmu a škody, které mohou vzniknout na lidském zdraví nebo na majetku. Jak tvrdí Ducháčková (2009), situace v podniku je hodnocena s ohledem na rizikovost, což především zahrnuje zjištění takových oblastí, jež jsou významné z pohledu realizace rizika. V této fázi jsou tedy obsažena zhodnocení:

- majetku nacházejícího se v podniku,
- umístění majetku (budovy, stavby, stroje apod.),
- finanční hodnoty (např. finanční hotovost, cenné papíry),
- pracovní síly,
- nehmotných aktiv (know-how, postavení na trhu).

Po zjištění stavu a situace podniku následuje určení rizik, která mohou mít negativní vliv na jednotlivé oblasti podniku. Pojišťovna může být ohrožena jednak riziky působícími uvnitř podniku (např. požár, poruchy zařízení, exploze, krádeže), ale také riziky ovlivňujícími její fungování zvnějšku (přírodní nebezpečí, politická rizika, ekologické problémy atp.). Tato rizika mohou být příčinou vzniku ztrát a škod:

- na osobách (např. nemoc, úraz, úmrtí),
- na majetku (např. krádež, požár),
- finančních (např. ztráta pozice na trhu, kurzové ztráty, odpovědnost za škody).

Shrneme-li výše uvedené, tak v rámci analýzy rizika se nejdříve zabýváme celkovou hospodářskou situací dané pojišťovny, poté jsou vymezeny druhy rizika, kterým je pojišťovna vystavována, případně možné expozice a celkový dopad na daný subjekt při realizaci daných rizik a pravděpodobnost, že tato rizika vůbec budou realizována.

### **2.3.2 OHODNOCENÍ A KVANTIFIKACE RIZIKA**

Hlavním smyslem této druhé fáze risk managementu je zjistit váhy, jaké náleží jednotlivým analyzovaným rizikům, a výši možných dopadů realizací rizik na finanční situaci dané pojišťovny. Stabilita určitého subjektu může být ovlivněna dopady ve formě materiálních (např. věcné škody, ztráta příjmů), ale i nemateriálních ztrát (např. ztráta dobré pověsti). Výsledky této fáze jsou následně podkladem pro rozhodování a přijímání opatření prováděné ve třetí fázi řízení rizik – kontrola a řízení rizika.

K měření rizika jsou v rámci risk managementu komerčními pojišťovnami zpravidla používány metody kvantitativní a metody kvalitativní. Kvantitativní metody jsou postaveny na matematickém výpočtu rizika z údajů o frekvenci výskytu hrozby a jejího případného dopadu. Výsledkem je nejčastěji výše roční předpokládané ztráty ve formě peněžního vyjádření. Tyto metody jsou velmi náročné na zpracování, proto pojistní analytici často pracují s počítačovými modely a dalšími speciálními programy. Naopak výstupem kvalitativních metod není žádná numerická hodnota, nýbrž charakteristika povahy rizika a seznam rizik zahrnující i popis jejich pravděpodobných výskytů. Díky těmto závěrům je pak rizikový manažer lépe schopen orientovat svůj čas na ty oblasti, jež jsou k riziku nejcitlivější, blíže viz Řezáč (2011).

V odvětví pojišťovnictví se obecně rozlišují dva základní typy rizik. Jedná se buď o rizika s vysokou četností, ale malou expozicí, anebo naopak o rizika s nízkou četností, ale vysokou expozicí. První zmiňovaný druh rizik se dá relativně dobře řídit, předvídat a také eliminovat, zatímco druhý jmenovaný typ rizik lze jen těžko odhadovat, jde totiž o tzv. katastrofická rizika, jejichž realizace je vysoce nepravděpodobná. Jak uvádí Ducháčková (2009), pro ohodnocení možných dopadů existence daných rizik na finanční situaci určitého pojistitele je potřeba zjistit pravděpodobnosti vzniku škody a možné odpovídající velikosti důsledků při realizaci rizika. Pojišťovny obvykle kalkulují s maximální možnou výší škody, která může realizací rizik nastat. V praxi se rozlišují následující přístupy pro stanovení předpokládané výše škody způsobené určitým rizikem:

- *MAS* (Maximum Amount Subject) – největší možná škoda, při výpočtu se vychází z nejnepríznivějších podmínek;

- *EML* (Estimated Maximum Loss) – odhadovaná největší škoda charakterizovaná jako maximální kombinovaná ztráta, přičemž nejsou uvažovány značně nepravděpodobné situace;
- *PML* (Probable Maximum Loss) – pravděpodobná největší škoda definovaná stejně jako *EML*, navíc se ale připouští působení lidského faktoru;
- *NLE* (Normal Loss Expectancy) – běžně očekávaná škoda specifikovaná jako *EML*, avšak se nadto počítá i s pohotovou lidskou reakcí.

### 2.3.3 KONTROLA A ŘÍZENÍ RIZIKA

V poslední ze tří fází risk managementu je zahrnut soubor několika činností, které spadají do této oblasti řízení. Hlavními úkoly kontroly a řízení rizika je přijímání preventivních opatření k předcházení škod, finanční eliminování rizika kvantifikovaného v předchozí fázi procesu risk managementu a následná kontrola a monitorování rizika. Každá pojišťovna se pak snaží činit na této úrovni řízení taková rozhodnutí, která by měla vést k zajištění stabilních a likvidních finančních toků, tedy k zabezpečení svého úspěšného a efektivního fungování.

Prvotním záměrem pojišťoven by mělo být konat kroky směřující k předcházení rizikům v co možná největší míře. Prevenci rizika lze provádět prostřednictvím zavádění strategických či fyzických opatření. Mezi strategická opatření kupříkladu můžeme zařadit užívání bezpečnějších technologií a materiálů či smluvní vyloučení odpovědnosti. Naopak do kategorie fyzických opatření lze zahrnout zavedení zařízení přímo chránícího před rizikem, jako například požární zařízení či bezpečnostní zámky, viz Ducháčková (2009).

Jak tvrdí Ducháčková, Daňhel (2010), risk management představuje u existujících rizik, kterým nelze předejít strategickými či fyzickými opatřeními, v podstatě rozhodování o tom, zdali se pojistitel vyrovná s riziky sám, nebo je předá dále ke krytí. Toto rozhodování o finanční eliminaci rizika je ovlivněno pravděpodobností výskytu rizika a intenzitou škody způsobené realizací příslušného rizika.

Jsou-li rizika kryta z vlastních zdrojů, využívá k tomu daný pojistitel svých běžných příjmů anebo tzv. samopojištění. Běžné příjmy jsou vhodným nástrojem ke snížení takových rizik, jejichž realizace se poměrně pravidelně opakuje a z hlediska výše expozice představuje malé ztráty. Nástroj samopojištění lze charakterizovat jako proces vytváření vlastních rezerv sloužících ke krytí určitých rizik, jehož výhodou jsou bezesporu nižší náklady, nicméně na druhou stranu jsou s ním spjaté problémy při vzniku velké až katastrofické škody nebo při vyčerpání rezervy z důvodu výskytu několika škod najednou, blíže viz Ducháčková (2009).

V případě, že se pojistitel rozhodne převést riziko na další subjekt, může tak učinit prostřednictvím tzv. tradičních způsobů eliminace rizik nebo netradičních alternativních nástrojů přenosu rizik. K tradičním nástrojům snižování rizika se řadí hlavně zajištění, které je spolu s dalšími možnostmi eliminace rizika stručně popsáno v podkapitole týkající se pojistně technických rezerv, soupojištění a pojistné pooly. Netradiční způsob alternativního přenosu rizik *ART* lze dle Cipra (2002, str. 172) definovat jako „*souhrnné označení pro různé metody, které představují vhodnější možnosti cedování pojistných rizik než klasické zajištění*“. Pod *ART* spadají především nástroje týkající se finitního zajištění, pojistných dluhopisů, pojistných derivátů či kontingentního kapitálu.

V neposlední řadě je velmi důležité v rámci risk managementu pravidelně sledovat a kontrolovat oblast rizik. Každá pojišťovna by měla mít vytvořený celkový systém přístupu k risk managementu, jehož efektivitu by měla neustále hodnotit a posuzovat. Dojde-li k výskytu nových rizik, je vhodné, aby je daná instituce co nejdříve zahrnula do svých modelů a dále s nimi pracovala.

## **2.4 REGULAČNÍ POJETÍ RISK MANAGEMENTU V POJIŠŤOVNICTVÍ**

Hlavním úkolem regulace v odvětví pojišťovnictví je zajistit bezpečnost a stabilitu komerčního pojistného systému, přičemž důraz je kladen především na ochranu uživatelů pojistných služeb, tedy klientů pojišťoven. Dalšími neméně důležitými cíli regulace pojišťovnictví jsou zabezpečení důvěryhodnosti daných finančních institucí a podpora transparentnosti v rámci provozování pojišťovacích služeb, viz Cipra (2002).

Na pojistném trhu působí specifické faktory, kterými je v podstatě vynucena regulace pojišťovací činnosti. Celou řadu různých důvodů pro regulaci odvětví pojišťovnictví zmiňují Ducháčková (2009) a Ducháčková, Daňhel (2010), nicméně dále jsou zaznamenány pouze ty nejvýznamnější faktory. První důležitou okolností je samotný charakter pojišťovacích služeb, kdy si klienti prostřednictvím předem zaplaceného pojistného kupují finanční krytí případných důsledků realizace vymezeného rizika, dochází zde tudíž k časovému nesouladu mezi placením pojistného a výplatou pojistného plnění. Vztah pojišťovny a klienta se tedy musí opírat o vzájemnou důvěru, a to hlavně ze strany pojistníka. Dále je pro pojistné odvětví typický systém tvorby rezerv disponující vysokým objemem peněžních prostředků, a proto je nezbytné striktně vymezit jednak povinnost tvořit určité druhy rezerv, ale rovněž způsob nakládání s těmito finančními prostředky. Na pojistných trzích je také v současné době nabízena rozsáhlá škála rozdílných pojistných produktů, pro jednotlivce je pak velmi obtížné se na trhu správně zorientovat a zvolit si nejlepší variantu vyhovující jeho potřebám. Zájemci

o pojišťovací služby jsou tak vystaveni hrozbě špatného výběru, kterou se snaží regulátoři snížit na minimum. Regulace a dohled se týká všech pojistitelů i zajišťitelů, kteří jsou oprávněni provozovat pojišťovací či zajišťovací činnost, ale také pojišťovacích zprostředkovatelů a dalších fyzických či právnických osob pohybujících se na pojistném trhu.

#### **2.4.1 LEGISLATIVNÍ RÁMEC RISK MANAGEMENTU**

Základním stavebním kamenem regulace pojišťovnictví je existence právních předpisů a norem upravujících tuto oblast finančních trhů. V České republice je dohled nad dodržováním právních regulí v pojišťovnictví vykonáván Českou národní bankou, a to od roku 2006. Snahou ČNB je především podporovat zdravý rozvoj, tržní disciplínu a konkurenceschopnost pojišťoven i zajišťoven na trhu, předcházet krizím systému, chránit pojistníky, pojištěné i oprávněné osoby a také posilovat důvěru veřejnosti v pojišťovnictví. Mezi základní pole působnosti ČNB jako orgánu dohledu se řadí povolovací, schvalovací, notifikační a kontrolní činnost.

Každá pojišťovna či zajišťovna sama nese zodpovědnost za funkčnost svých vnitřních kontrolních mechanismů a systému řízení rizik. Činnost pojišťoven a zajišťoven je upravena zejména zákonem č. 277/2009 Sb., o pojišťovnictví a příslušnými prováděcími vyhláškami. Jak uvádí Ducháčková, Daňhel (2010), tento zákon o pojišťovnictví je vytvořen v souladu s předpisy Evropských společenství a jsou v něm obsaženy významné věcné změny v oblasti regulace pojišťovnictví. Od nabytí jeho účinnosti je po komerčních pojišťovnách vyžadováno vytvoření a udržení řídicího a kontrolního systému jako mezistupně k implementaci konceptu Solvency II. Jak sekce risk managementu, tak vnitřní kontrolní systém pojišťovny se musí maximálně snažit předcházet negativním ekonomickým výsledkům v důsledku realizace všech podstupovaných rizik. Tímto zákonem jsou dále kupříkladu zpřísněny podmínky týkající se uzavírání doplňkových pojištění a přepracována je také oblast výpočtů a použití technických rezerv, kvótování aktiv a možnosti investiční politiky pojišťovny.

Dle Ducháčková (2009) je Česká republika členským státem Evropské unie, a tudíž je zároveň součástí Jednotného evropského pojišťovacího trhu a musí vycházet při regulaci pojišťovnictví z legislativních předpisů Evropské unie. V rámci společného trhu jsou tedy zákonem vymezena základní pravidla pro regulaci odvětví pojišťovnictví, která by měla být východiskem pro právní úpravy provedené jednotlivými členskými státy. Vývoj legislativního rámce pojistného trhu pro potřeby Evropské unie lze označit jako velmi dlouhodobý. Nejdříve byly vypracovány směrnice první generace (v r. 1973), na které postupně navázaly druhá (v r. 1988) a třetí generace směrnic (v r. 1992). Zákonodárci Evropské unie také v souladu s třetí



generací směrnic vytvořili regulační koncept řízení rizik pro oblast pojišťovnictví, jež je pojmenován jako Solvency I, respektive Solvency II přiblížený v následující kapitole.

Závěrem této části lze konstatovat, že v oblasti pojišťovnictví dochází v posledních letech k upřednostňování regulace finanční stability jednotlivých institucí a zajištění vyšší míry splnitelnosti závazků pojistiteli. Regulační předpisy se v nedávné době začaly orientovat také na sledování odpovídající tvorby technických rezerv a solventnosti pojišťoven. Dalším přínosem regulace a vykonávaného dohledu je zvýšení míry transparentnosti při poskytování pojistných produktů projevující se zejména ve vyšší informovanosti klientů o jednotlivých pojistných produktech.

#### **2.4.2 KONCEPT SOLVENCY II**

Metodika Solvency II je spojena s přijetím Směrnice Evropského parlamentu a Rady 2009/138/ES ze dne 25. listopadu 2009, o přístupu k pojišťovací a zajišťovací činnosti a jejím výkonu, a Směrnice 2014/51/EU Evropského parlamentu a Rady ze dne 16. dubna 2014, kterou se mění směrnice 2009/138/ES a další nařízení s ohledem na pravomoci Evropského orgánu dohledu. Uvedené změny vztahené k Solvency II měly nabýt účinnosti dne 1. ledna 2016, nicméně Poslanecká sněmovna České republiky návrh novely zákona o pojišťovnictví neschválila. Ministerstvo financí České republiky tak připravilo transpoziční novelu zákona o pojišťovnictví zohledňující jen požadavky vyplývající ze směrnice Solvency II, jež byla předložena Poslanecké sněmovně naposledy dne 3. února 2016, jak se uvádí na internetových stránkách ministerstva. Lze očekávat, že Ministerstvo financí tuto novelu opět v nejbližších dnech předloží sněmovně ke schválení, takže se můžeme jen domnívat, kdy skutečně vejdou tyto změny v platnost.

Stěžejním pojmem celé metodiky Solvency II je *solventnost pojišťovny*, což lze vymezit dle Cipra (2002) jako schopnost pojistitele plnit přijaté pojistné závazky neboli jinak řečeno včas a ve stanoveném objemu uhradit oprávněné pojistné nároky z realizovaných pojistných událostí. Solventnost pojišťovny můžeme tedy chápat jako dostupnost volných a vlastních kapitálových prostředků k plnění současných i budoucích závazků. Solventnosti lze dosáhnout především dostatečnou kapitálovou vybaveností a odpovídající výší a kvalitou technických rezerv pojistitele. Právě oblast solventnosti je hlavním předmětem zájmu orgánů regulace.

Jak vyplývá z názvu, systém Solvency II je již druhou metodikou připravenou orgány Evropské unie k začlenění do právních systémů jednotlivých členských států. Úkolem konceptu Solvency II je komplexnější zachycení finančního zdraví a solventnosti pojišťoven a také sjednocení regulačního režimu v rámci Evropské unie. Hlavní změnou oproti Solvency I

je zohledňování rizikového profilu daných pojistitelů a snaha o motivaci institucí pro zkvalitnění systému řízení rizik.

Jak tvrdí Ducháčková, Daňhel (2010), koncept Solvency II je tvořen třemi základními pilíři opírajícími se o:

- pravidla týkající se finančních zdrojů pojišťoven, tedy velikosti kapitálu (minimální výše kapitálu, rizikový kapitál a technické rezervy);
- pravidla určená pro vytváření vnitřních modelů a postupů pro risk management;
- pravidla směřující k vyšší transparentnosti, navíc s významnou rolí ratingových agentur.

Výše zmíněné tři pilíře systému Solvency II jsou znázorněny pomocí Obr. 2.1 níže.

*Obr. 2.1: Schéma třípilířové struktury Solvency II*

<b>I. PILÍŘ</b>	<b>II. PILÍŘ</b>	<b>III. PILÍŘ</b>
<b>KVANTITATIVNÍ POŽADAVKY</b>	<b>KVALITATIVNÍ POŽADAVKY</b>	<b>TRŽNÍ DISCIPLÍNA</b>
<ul style="list-style-type: none"> <li>▪ Základní principy pro kalkulace technických rezerv</li> <li>▪ Minimální kapitál</li> </ul>	<ul style="list-style-type: none"> <li>▪ Pravidla interního řízení rizik</li> <li>▪ Proces dohledu</li> </ul>	<ul style="list-style-type: none"> <li>▪ Zveřejňování informací o rizicích a scénářích</li> <li>▪ Transparentnost</li> </ul>

*Zdroj: Ducháčková, Daňhel (2010, str. 101)*

*První pilíř* konceptu Solvency II v sobě zahrnuje regulaci technických rezerv, určení pravidel pro investování pojistitelů do jednotlivých investičních instrumentů, zjištění solventnostního kapitálového požadavku a stanovení minimálního požadavku na kapitál představujícího určitou garanční pojistku.

*Druhý pilíř* Solvency II je zaměřen hlavně na kvalitativní aspekt řízení rizik pojišťovnou. V rámci této sekce se provádí sledování systémů interní kontroly a řízení rizik, hodnocení a řízení rizik z kvalitativního hlediska a dohled regulátora zabývající se hlavně funkčností systémů řízení pojišťovny.

*Třetí pilíř* metodiky Solvency II je orientován především na tržní disciplínu, pod čímž si lze představit povinné zveřejňování informací jednotlivými pojišťovnami, zvyšování transparentnosti činnosti pojistitelů a celkové posílení srovnatelnosti na trhu nabízených pojistných produktů, blíže viz Ducháčková, Daňhel (2010).

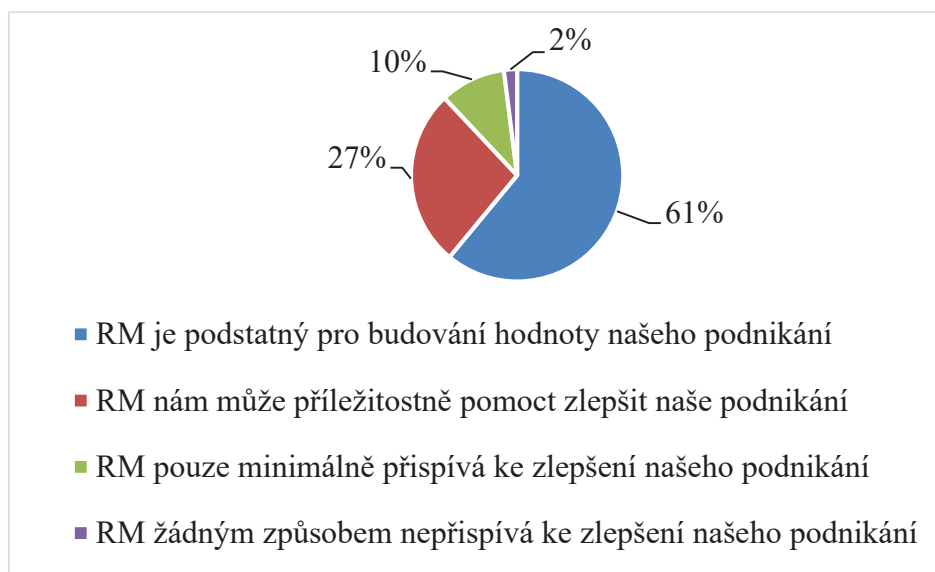


## 2.5 AKTUÁLNÍ STAV RISK MANAGEMENTU V POJIŠŤOVNICTVÍ

Propuknutí celosvětové finanční krize významným způsobem ovlivnilo odvětví pojišťovnictví, i když zde došlo k relativně menším finančním ztrátám oproti bankovnímu sektoru. Zejména v té době se ukázala nutnost zpřísnit regulaci všech finančních trhů, pojišťný trh nevyjímaje. Jelikož systém Solvency I, podobně jako Basel upravující sektor bankovní, nevykazoval žádné příznaky nerovnováhy signalizující příchod krize, bylo potřeba pokračovat v implementaci a navrhnout nový přístup řešení v podobě Solvency II, díky kterému se dostala do popředí zájmu pojišťoven činnost komplexního řízení rizik.

Jednotliví pojistitelé v posledních letech stále více zaměřují svou pozornost na kvalitní risk management podporovaný sofistikovaným modelováním rizik společně s analyzováním možných scénářů, a připravují se tak na nabytí účinnosti nového regulačního rámce evropského pojišťovnictví – Solvency II. V jednotlivých pojišťovnách vznikají sekce tzv. enterprise risk managementu (ERM), pod čímž si lze představit integrované řízení firemních rizik ohrožujících dosažení strategických cílů daného pojistitele. Větší orientace pojišťoven na risk management je deklarována také vzrůstající rolí ředitele oddělení řízení rizik v rámci celé výkonné rady daného podnikatelského subjektu. Výzkumem společnosti KPMG provedeném v roce 2013 a zaznamenaném v Graf 2.1 se dále poukazuje na fakt, že i samotní pojistitelé již nepohlíží na koncepci řízení rizik pouze jako na činnost vyplývající z příslušných právních regulí, ale hlavně jako na činnost výrazně přispívající k úspěšnosti podnikání.

Graf 2.1: Hodnota risk managementu vnímaná pojistiteli



Zdroj: <https://www.kpmg.com/PE/es/IssuesAndInsights/ArticlesPublications/Documents/Insurance-risk-and-capital-transformation.pdf>

Je zřejmé, že významný progres v přístupu pojišťoven k risk managementu byl spuštěn regulatorními předpisy, nicméně můžeme také konstatovat, že dle Graf 2.1 si již nejméně dvě třetiny pojistitelů uvědomují přínos risk managementu jako součásti úspěšného a efektivního řízení podniku.

Ze závěrů kvantitativního výzkumu zachyceného v publikaci Řezáč (2011) a zpracovaného v roce 2011 lze odvodit následující skutečnosti týkající se řízení rizik ve vybraných pojišťovnách:

- risk managementu je věnována rozdílná pozornost jednotlivými pojistiteli;
- v důsledku finanční krize byla pojišťovnami restrukturalizována rizika;
- v souladu s konceptem Solvency II je pozornost pojišťoven soustředěna na kapitálovou vybavenost a kvalitu managementu;
- snižuje se zájem o řízení operačního rizika;
- častěji jsou využívány opracované matematicko-statistické modely;
- risk management by stále měl být jedním z hlavních zájmů managementu pojišťoven.

Jelikož v posledních letech také neustále přibývají rizika, jež mohou mít výrazný negativní dopad na hospodaření pojišťoven (např. nárůst hrozby terorismu, přírodních katastrof, internetových útoků), je nezbytné, aby pojistitelé dále pokračovali v nastaveném trendu a neustále se zabývali možnými zlepšeními na úrovni řízení rizik a kapitálu. Vhodné by bylo, aby jednotlivé pojišťovny preferovaly vlastní modely lépe odrážející jejich rizikový profil před standardními modely vytvořenými příslušnými regulátory.

### 3 CHARAKTERISTIKA POUŽITÉ METODIKY

V této kapitole se blíže zaměříme na možnosti, kterými lze modelovat výši pojistné škody spojené s danou smlouvou. Modelů, které můžeme využít, existuje celá řada, nicméně v oblasti pojistných škod je vhodné pracovat s regresním modelem ze skupiny zobecněných lineárních modelů, je-li předpokladem gamma rozdělení. Dále lze k tomuto problému přistoupit z pohledu zohlednění extrémně velkých či extrémně malých ztrát, a to pomocí aplikace teorie extrémních hodnot, konkrétně metody „Peak over the threshold“, kdy se snažíme sestavit zobecněný lineární model za předpokladu smíšeného rozdělení zahrnujícího jak gamma rozdělení, tak také zobecněné Paretovo rozdělení pravděpodobnosti. Kromě charakteristiky těchto přístupů je obsahem kapitoly vymezení metody maximální věrohodnosti, dále principy parametrického Z-testu a popis analýzy reziduí. Ve třetí kapitole je čerpáno hlavně z odborné literatury Beirlant, Goegebeur, Segers, Teugels (2004), Coles (2001), Hardin, Hilbe (2007), Jong, Heller (2008) a Ohlsson, Johansson (2010).

#### 3.1 ZOBECNĚNÉ LINEÁRNÍ MODELY

Teorie zobecněných lineárních regresních modelů (*GLM*) byla poprvé představena v roce 1972 pány Nelderem a Wedderburnem, kteří vytvořili poměrně mocný statistický nástroj pro analyzování vícerozměrných datových souborů a vztahů mezi jednotlivými proměnnými. Zobecněné lineární modely se začaly více používat v oblasti pojistného trhu po roce 1990, v současné době jsou hojně uplatňovány kromě pojišťovnictví a oblasti financí i v dalších oborech, např. v odvětví medicíny či biologie.

Jak uvádí Hardin, Hilbe (2007), vytvořením konceptu *GLM* v podstatě došlo k rozšíření klasických lineárních regresních modelů. Konkrétně autoři konceptu zjistili, že obecné modely se mohou dále rozvíjet uvolněním některých předpokladů lineárního regresního modelu a že je v podstatě možné „linearizovat“ vztahy mezi proměnnými, které se původně jeví jako nelineární.

Pomocí zobecněných lineárních modelů je tedy možné vysvětlit variabilitu jedné náhodné proměnné změnami jiných dalších proměnných. Vysvětlovanou proměnnou lze označit pojmem závislá proměnná, zatímco vysvětlující proměnné jsou nezávislými veličinami. Tyto nezávislé proměnné jsou pak v rámci této teorie zejména v oblastech finanční a pojišťovnictví velmi často nazývány jako tzv. *rizikové faktory* (risk factors), blíže viz Jong, Heller (2008).

V rámci finančního modelování lze náhodnou veličinu chápat jako kombinaci systematické a náhodné složky. Prostřednictvím zobecněných lineárních modelů je pak možné

na základě historických údajů předpovídat systematickou složku při respektování náhodnosti podkladového děje. V teorii *GLM* je zahrnut obsáhlý soubor statistických metod a modelů, které zobecňují lineární regresní model těmito předpoklady:

- rozdělení náhodné složky  $y$  spadá do tzv. exponenciální rodiny rozdělení pravděpodobnosti s hustotou

$$f(y; \theta, \phi) = \exp \left\{ \frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi) \right\}, \quad (3.1)$$

kde symbolem  $y$  značíme vysvětlovanou proměnnou,  $\theta$  představuje kanonický parametr neboli link funkci,  $b(\theta)$  je kumulativní funkce,  $\phi$  je označením pro neznámý disperzní parametr a  $c(y, \phi)$  je tzv. normalizační člen;

- systematická složka je kombinací vektoru odhadovaných parametrů  $\beta$  a matice známých regresorů  $X$  a odpovídá lineárnímu prediktoru  $\eta$  ve tvaru

$$\eta = X \cdot \beta; \quad (3.2)$$

- známá monotónní diferencovatelná funkce vztahující lineární prediktor ke střední hodnotě  $\mu$  je označována jako link funkce  $g$  a lze ji zapsat pomocí rovnice

$$E[y] = \mu = g^{-1}(\eta); \quad (3.3)$$

- rozptyl náhodné veličiny  $V(y)$  je funkcí střední hodnoty a odvíjí se od typu exponenciálního rozdělení dle vztahu

$$V(y) = a(\phi) \cdot V(\mu). \quad (3.4)$$

Mezi základní typy exponenciálního rozdělení náhodné vysvětlované veličiny se řadí normální, binomické, inverzní Gaussovo, Poissonovo, gamma, geometrické či negativní binomické rozdělení. Každé toto rozdělení je pak určeno příslušnou střední hodnotou a rozptylem. Skupina exponenciálních rozdělení v podstatě poskytuje možnost modelovat spojitě, diskrétní, proporcionální, kategorické, ale také binární veličiny.

Jelikož v rámci zobecněných lineárních modelů jsou zkoumány vztahy mezi vysvětlovanou proměnnou s rozdělením z exponenciální skupiny rozdělení a celou řadou rizikových faktorů, je potřeba provést výběr statisticky významných regresorů. Tento problém může být vyřešen statistickou verifikací - konkrétně jednorozměrnou analýzou na bázi *Z*-testu. Po zahrnutí statisticky významných faktorů do regresního modelu lze získat odhady parametrů i v případě složitě formulovaných modelů. Výsledkem modelování pak mohou být buď bodové, nebo intervalové odhady, nicméně v praktické části práce se zabýváme pouze

odhady bodovými. V praxi jsou aplikovány různé metody bodového odhadu parametrů, přičemž mezi ty tradiční lze zařadit metodu momentů či metodu maximální věrohodnosti, která je použita i v této diplomové práci. Nyní je potřeba si uvědomit, že v dalším textu práce je vhodné nejdříve samostatně vymezit metodu maximální věrohodnosti a statistický Z-test tak, abychom pochopili jejich základní principy, a až poté přistoupit zpět k formulaci regresního modelu.

### 3.2 METODA MAXIMÁLNÍ VĚROHODNOSTI

Pomocí metody maximální věrohodnosti (*MLM*) lze získat maximálně věrohodné odhady neznámých parametrů modelu souhrnně značených symbolem  $\theta$ , a to pro různé formy rozdělení pravděpodobnosti. Snahou *ML* metody je nalézt hodnotu, na jejíž úrovni je dosaženo maximální pravděpodobnosti předpokladu, že rozdělení zkoumaných hodnot odpovídá očekávanému rozdělení. Jak uvádí Hardin, Hilbe (2007), hlavním principem této metody je tedy maximalizace tzv. věrohodnostní funkce, která se odvíjí od tvaru předpokládané funkce hustoty pravděpodobnosti pozorovaných hodnot  $y_i$ . Obecně lze tedy věrohodnostní funkci  $L$  zapsat pomocí rovnice

$$L(y; \theta, \phi) = \prod_{i=1}^n f(y_i; \theta, \phi), \quad (3.5)$$

kde je symbolem  $\theta$  míněn vektor neznámých odhadovaných parametrů vyjádřených v další části práce prostřednictvím link funkce s  $\beta$  parametry a  $\phi$  je označením pro neznámý odhadovaný parametr disperze.

Jelikož je těžké početně získat hodnotu věrohodnostní funkce formulované rovnicí (3.5), je tento vztah obvykle upravován prostřednictvím logaritmické transformace do následujícího tvaru

$$l(y; \theta, \phi) = \sum_{i=1}^n \log f(y_i; \theta, \phi), \quad (3.6)$$

ve kterém je zahrnuta právě logaritmická věrohodnostní funkce  $l$ . Tato transformace nemá žádný vliv na hodnoty extrémů věrohodnostní funkce, a proto se provádí téměř vždy při aplikaci *MLM*.

Dosadíme-li si do rovnice (3.6) funkci hustoty pravděpodobnosti pro skupinu exponenciálních rozdělení formulovanou matematickým vztahem (3.1), získáme po logaritmické transformaci funkci věrohodnosti v podobě

$$l(y; \theta, \phi) = \sum_{i=1}^n \left\{ \frac{y_i \theta_i - b(\theta_i)}{a(\phi)} + c(y_i, \phi) \right\}. \quad (3.7)$$

S logaritmickou věrohodnostní funkcí lze dále pracovat a upravovat ji, a to hlavně prostřednictvím první či druhé parciální derivace. Předpokládáme-li, že střední hodnotu  $\mu$  lze vyjádřit prostřednictvím link funkce nahrazující odhadovaný parametr  $\theta$ , lze po provedení první parciální derivace věrohodnostní funkce  $l$  dle parametru  $\beta$  získat tzv. gradient. Pokud tuto derivaci položíme rovnu 0, měli bychom dosáhnout odhadu parametrů  $\beta$ , a to dle vztahu

$$l' = \frac{\partial l}{\partial \beta} = 0. \quad (3.8)$$

Druhou parciální derivací funkce  $l$  vypočítáme tzv. *Hessian* neboli matici druhých derivací ( $H$ ), kterou lze zapsat v podobě

$$l'' = \frac{\partial^2 l}{\partial \beta \partial \beta^T} = H. \quad (3.9)$$

Výpočet matice  $H$  je velmi důležitý, neboť na diagonále matice inverzní k matici  $H$  se nachází druhé mocniny standardních chyb odhadu. Tato inverzní matice se nazývá informační matice a v této diplomové práci jsou její diagonální prvky použity k výpočtu  $Z$ -statistiky.

Pro maximalizaci věrohodnostní funkce u *GLM* se využívají iterační metody, nejčastěji algoritmus iterativních vážených nejmenších čtverců či Newtonův-Raphsonův algoritmus. Výsledný odhad parametrů by měl být konzistentní, eficientní a asymptoticky normální.

### 3.3 JEDNOROZMĚRNÁ ANALÝZA POMOCÍ Z-TESTU

Jednorozměrná analýza je velmi často používána k ověření hodnot parametrů základního souboru s určitým předpokládaným typem rozložení pravděpodobnosti, kdy se testuje zejména střední hodnota a rozptyl v porovnání s konkrétní očekávanou hodnotou. Tuto metodu analýzy dat je ale také možné uplatnit pro zjištění statistické významnosti odhadnutých parametrů modelu. Pro výběr významných regresních faktorů je zvolen právě tento statistický nástroj, díky kterému dojde k vyčlenění statisticky nevýznamných faktorů z celkového modelu.

Z existujících jednovýběrových parametrických testů je v této práci uplatněn  $Z$ -test, který je určen pro soubory s velkým rozsahem dat. Východiskem pro rozhodnutí, zdali danou proměnnou zahrnout do modelu, může být hodnota  $Z$ -statistiky nebo  $p$ -hodnota, jejichž výpočet je společně s postupem  $Z$ -testu přiblížen v následujících odstavcích.

Prvním krokem každého statistického testu je formulace hypotéz na předem stanovené hladině významnosti  $\alpha$ . V našem případě lze hypotézy nastavit takto:

$$H_0: \beta_i = 0; \quad (3.10)$$

$$H_A: \beta_i \neq 0. \quad (3.11)$$

Potvrdí-li se výchozí nulová hypotéza, pak daný parametr, ani příslušný regresní faktor, není statisticky významný a nezahrne se do celkového modelu. V případě přijetí alternativní hypotézy se daný faktor stane součástí regresního modelu.

Dále je nezbytné vyčíslit hodnotu zvoleného testovacího kritéria – Z-statistiky, a to dle vzorce

$$Z = \frac{\hat{\beta}_i - \beta_0}{SE_{\hat{\beta}_i}}, \quad (3.12)$$

kde  $\hat{\beta}_i$  představuje odhad parametru  $\beta$  metodou maximální věrohodnosti,  $\beta_0$  je označením pro hodnotu parametru v případě nulové hypotézy a  $SE$  je chyba příslušného odhadu, jejíž hodnotu můžeme získat z diagonály informační matice zmíněné v předchozí kapitole.

Dalším krokem je určení dolní a horní hranice konfidenčních intervalů, které lze v případě 95% hladiny spolehlivosti stanovit jako

$$LCI = \hat{\beta}_i - 1,96 \cdot SE_{\hat{\beta}_i} \quad \wedge \quad UCI = \hat{\beta}_i + 1,96 \cdot SE_{\hat{\beta}_i}, \quad (3.13)$$

kde  $LCI$  představuje spodní hranici a  $UCI$  horní hranici konfidenčního intervalu. Pokud se testovaná Z-statistika nachází uvnitř kritického oboru  $(-\infty; LCI) \cup (UCI; \infty)$ , zamítáme nulovou hypotézu ve prospěch alternativní hypotézy na dané hladině významnosti  $\alpha$ , odhadovaný parametr je tedy spolu s regresním faktorem statisticky významný. V opačném případě nulovou hypotézu nezamítáme a s daným faktorem již dále nepracujeme.

K Z-testu lze přistoupit také z pohledu  $p$ -hodnoty, přičemž postup odpovídá výše popsanému až do bodu výpočtu Z-statistiky. Hodnota tohoto testovacího kritéria se musí převést do pravděpodobnostní škály na  $p$ -hodnotu, což se obvykle provádí prostřednictvím funkcí obsažených v programu MS Excel. Výsledná  $p$ -hodnota se pak srovnává s předem nastavenou hladinou významnosti. Pokud je  $p$ -hodnota nižší než určená hladina významnosti, zamítáme nulovou hypotézu a přijímáme hypotézu alternativní, daný parametr je tedy statisticky významný. V případě, že je  $p$ -hodnota vyšší než stanovená hladina významnosti, je přijata nulová hypotéza.

### 3.4 REGRESNÍ GAMMA MODEL

Při konstrukci zobecněného lineárního modelu je vhodné postupovat dle následujících kroků. Nejdříve je potřeba určit typ rozdělení pravděpodobnosti náhodné veličiny  $y_i$ , v našem případě velikosti pojistné škody. Dále je žádoucí zvolit si takovou link funkci, jež bude co nejlépe odpovídat vývoji střední hodnoty daného rozdělení. Po provedení výběru statisticky významných regresních faktorů dle již dříve popsaného Z-testu jsou poté pomocí metody

maximální věrohodnosti odhadovány parametry celkového regresního modelu. Při tvorbě gamma modelu se obvykle pracuje s dvěma parametry gamma rozdělení, přičemž v této práci se odhadují konstantní hodnota variačního parametru a na střední hodnotu navázané parametry příslušné link funkce.

### 3.4.1 GAMMA ROZDĚLENÍ PRAVDĚPODOBNOSTI

Hlavním zájmem praktické části diplomové práce je modelování výše pojistné škody, tudíž je potřeba najít pro tuto náhodnou veličinu vhodné rozdělení. Jak uvádí Ohlsson, Johansson (2010), gamma rozdělení se pro modelování pojistných škod stalo postupem času v podstatě standardně aplikovaným typem rozdělení, a to i přestože existují další rozdělení splňující základní požadavky pozitivního a doprava zešikmeného rozdělení. Toto rozdělení pravděpodobnosti spadá do kategorie spojitých pravděpodobnostních rozdělení, jež se obvykle vyjadřují pomocí funkce hustoty pravděpodobnosti a kumulativní distribuční funkce. Náhodné veličiny  $y_i$  z gamma rozdělení nabývají pouze kladných hodnot, stejně jako parametry tohoto rozdělení, které tedy musí být pozitivními reálnými čísly.

Chceme-li formulovat funkci hustoty pravděpodobnosti  $f$  pro gamma rozdělení, případně i kumulativní distribuční funkci, je možné využít různé parametrizace zmiňované v odborných publikacích, avšak nejčastěji se uvádí tři typy definované v následujících odstavcích. V této diplomové práci je pak použita poslední specifikovaná parametrizace.

V rámci první parametrizace se pracuje s parametrem tvaru  $k$  (tzv. shape parametr) a parametrem variability označovaným symbolem  $\theta$  (tzv. scale parametr). Funkci hustoty pravděpodobnosti lze v tomto případě zapsat jako

$$f(y; k, \theta) = \frac{y^{k-1} \cdot e^{-\frac{y}{\theta}}}{\Gamma(k) \cdot \theta^k}. \quad (3.14)$$

Dále je možné funkci hustoty pravděpodobnosti  $f$  vyjádřit zahrnutím parametrů  $\alpha$  (shape parametr) a  $\beta$  (inverzní parametr variability), tedy pomocí matematického vztahu

$$f(y; \alpha, \beta) = \frac{y^{\alpha-1} \cdot e^{-y \cdot \beta} \cdot \beta^\alpha}{\Gamma(\alpha)}, \quad (3.15)$$

kde

$$\alpha = k \quad \wedge \quad \beta = \frac{1}{\theta}. \quad (3.16)$$

Posledním a nejčastěji v praxi využívaným přepisem funkce hustoty pravděpodobnosti  $f$  je rovnice zachycující variační parametr  $\phi$  a parametr střední hodnoty  $\mu$  v podobě



$$f(y; \mu, \phi) = \frac{1}{y \cdot \Gamma\left(\frac{1}{\phi}\right)} \cdot \left(\frac{y}{\mu \cdot \phi}\right)^{\frac{1}{\phi}} \cdot e^{-\frac{y}{\mu \cdot \phi}}, \quad (3.17)$$

kde

$$\frac{1}{\phi} = k = \alpha \quad \wedge \quad \mu = \frac{k}{\beta} = k \cdot \theta. \quad (3.18)$$

### 3.4.2 ODVOZENÍ GAMMA MODELU

Základní funkce hustoty pravděpodobnosti pro gamma rozdělení je zachycena pomocí rovnice (3.17) výše. Bereme-li v úvahu, že gamma rozdělení je exponenciálního typu, lze tento vztah přepsat do tvaru

$$f(y; \mu, \phi) = \exp\left\{\frac{y/\mu - (-\ln \mu)}{-\phi} + \frac{1-\phi}{\phi} \cdot \ln y - \frac{\ln \phi}{\phi} - \ln \Gamma\left(\frac{1}{\phi}\right)\right\}, \quad (3.19)$$

kde symbolem  $y$  je označena závislá proměnná, v našem případě výše pojistné škody,  $\mu$  představuje parametr střední hodnoty a  $\phi$  parametr disperze.

Jak zmiňuje Hardin, Hilbe (2007), z rovnice (3.19) můžeme zjistit, že funkci  $b(\theta)$  nazývanou jako kumulativní funkce lze zapsat v podobě

$$b(\theta) = -\ln \mu \quad (3.20)$$

a samotnou link funkci  $\theta$  ve tvaru

$$\theta = \frac{1}{\mu}. \quad (3.21)$$

Z kumulativní funkce lze odvodit první a druhou parciální derivací funkci střední hodnoty a funkci rozptylu  $V(\mu)$ . První parciální derivací je tedy střední hodnota získána postupem

$$b'(\theta) = \frac{\partial b}{\partial \mu} \frac{\partial \mu}{\partial \theta} = \quad (3.22)$$

$$= \left(-\frac{1}{\mu}\right) \cdot (-\mu^2) = \quad (3.23)$$

$$= \mu. \quad (3.24)$$

Pomocí druhé derivace můžeme zjistit funkci rozptylu následovně

$$b''(\theta) = \frac{\partial^2 b}{\partial \mu^2} \left(\frac{\partial \mu}{\partial \theta}\right) + \frac{\partial b}{\partial \mu} \frac{\partial^2 \mu}{\partial \theta^2} = \quad (3.25)$$

$$= (1)(-\mu^2) = \quad (3.26)$$

$$= -\mu^2. \quad (3.27)$$

Tento rozptyl střední hodnoty označovaný jako  $V(\mu)$  je součástí rozptylu náhodné veličiny  $V(y)$ , jejich vzájemný vztah je obsažen již v rovnici (3.4). Funkci  $V(y)$  lze ale také vyjádřit jako součin  $b''(\theta)$  a  $a(\phi)$ , z čehož lze dále usuzovat, že rozptyl  $V(\mu)$  může být prezentován i ve tvaru

$$V(\mu) = \mu^2. \quad (3.28)$$

### 3.4.3 ODHAD PARAMETRŮ GAMMA MODELU

Jak již bylo uvedeno výše, pro odhad parametrů regresního modelu pracujeme v této diplomové práci s metodou maximální věrohodnosti samostatně charakterizovanou v kapitole 3.2. Na základě principu *MLM* lze zlogaritmovanou věrohodnostní funkci pro gamma model zapsat pomocí vzorce

$$l(y; \mu, \phi) = \sum_{i=1}^n \left\{ \frac{y_i / \mu_i - (-\ln \mu_i)}{-\phi} + \frac{1-\phi}{\phi} \cdot \ln y_i - \frac{\ln \phi}{\phi} - \ln \Gamma\left(\frac{1}{\phi}\right) \right\}, \quad (3.29)$$

jež se může ještě dále upravit podle typu zvolené link funkce. Standardně se v případě gamma rozdělení používá kanonická link funkce v podobě

$$g^{-1}(\eta) = \mu = \frac{1}{X \cdot \beta}, \quad (3.30)$$

která se následně dosazuje za parametr střední hodnoty do vzorce (3.29). Pro účely této diplomové práce nicméně bereme v úvahu nekanonickou link funkci pro střední hodnotu určenou pomocí vztahu

$$g^{-1}(\eta) = \mu = \exp(X \cdot \beta) \quad (3.31)$$

a nazývanou jako log-link funkce. Věrohodnostní funkci po logaritmické transformaci lze pak pro regresní model s log-link funkcí stanovit jako

$$l = \sum_{i=1}^n \left\{ \frac{y_i / \exp(x_i \cdot \beta) + x_i \cdot \beta}{-\phi} + \frac{1-\phi}{\phi} \cdot \ln y_i - \frac{\ln \phi}{\phi} - \ln \Gamma\left(\frac{1}{\phi}\right) \right\}. \quad (3.32)$$

S takto odvozeným regresním modelem pracujeme v praktické části práce. Model můžeme v podstatě nazývat rovněž jako *log-gamma model*, neboť je v něm zahrnuta log-link funkce. Důvodem pro výběr této varianty ze skupiny gamma modelů je především lehčí interpretace a také snadnější formulace omezujících podmínek pro hodnotu odhadovaných parametrů  $\mu$  a  $\phi$ , respektive  $\beta$  a  $\phi$ . Pomocí log-gamma modelu je tedy možné modelovat pojistné škody na bázi gamma rozdělení, jelikož výpočetní algoritmus je stejný jako u klasického gamma modelu s kanonickou link funkcí. Log-gamma model se avšak oproti

klasickému gamma modelu liší ve skutečnosti, že regresí jsou získány jiné standardní chyby odhadu ( $SE$ ), nicméně jejich vzájemný rozdíl je, až na extrémní případy, minimální, blíže viz Hardin, Hilbe (2007).

K vyhodnocení přesnosti odhadnutého regresního modelu lze použít různé statistické nástroje, v praktické části diplomové práce je aplikována reziduální analýza, konkrétně jsou zjišťovány hodnoty tzv. hrubých reziduí a Pearsonových reziduí. Analýze reziduí se blíže věnujeme samostatně v závěrečné kapitole třetí části práce.

### 3.5 TEORIE EXTRÉMních HODNOT

Teorie extrémních hodnot ( $EVT$ ) neboli také teorie extrémních ztrát je dynamicky se rozvíjející oblastí matematické statistiky, která slouží k modelování výběrových extrémních hodnot stejně rozdělené náhodné veličiny, tedy jejích maxim nebo minim. Tento přístup je možné uplatnit v rozličných sférách a oborech, zejména je využíván v hydrologii, meteorologii, metalurgii a v posledních letech také stále více ve finanční oblasti a v pojišťovnictví. V případě posledního zmiňovaného odvětví je hlavním polem aplikace  $EVT$  oblast neživotního pojištění, kde výskyt extrémní pojistné škody může velmi ohrozit solventnost daného pojistitele. Z tohoto důvodu se pozornost pojistných analytiků upíná na vytváření modelů zachycujících extrémní hodnoty a na následné vhodné sestavení jednotlivých pojistných kmenů.

Jak uvádí Coles (2001), základem teorie extrémních hodnot je zkoumání statistických vlastností množiny  $M_n$ , ve které jsou obsažena maxima posloupnosti nezávislých náhodných proměnných  $y_i$  majících společnou distribuční funkci  $F$ . Pozornost je zaměřena zejména na pravděpodobnostní chování maxima  $M_n$  pro vysoké hodnoty  $n$ , což v podstatě představuje analyzování limitního vztahu náhodné veličiny  $M_n$  pro  $n$  jdoucí k nekonečnu. V případě, že uvažujeme  $y^*$  jako pravý koncový bod rozdělení určeného distribuční funkcí  $F$ , pak je platný zápis

$$M_n \rightarrow y^*, \quad (3.33)$$

a rozdělení množiny  $M_n$  tedy konverguje ke konstantě na úrovni pravého koncového bodu. Provedeme-li úpravu proměnných této množiny prostřednictvím lineární normalizace, získáme množinu  $M_n^*$  ve tvaru

$$M_n^* = \frac{M_n - b_n}{a_n}, \quad (3.34)$$

kde  $a_n > 0$  a  $b_n$  jsou posloupnosti reálných konstant. Pokud jsou nalezeny vhodné normalizační množiny  $a_n$  a  $b_n$ , lze takové limitní rozdělení označit za nedegenerované a

můžeme jej specifikovat distribuční funkcí  $G(y)$ , přičemž významnou výhodou zavedené normalizace je stabilita stejně rozdělených náhodných veličin  $y_i$  v maximu.

Zásadním východiskem pro teorii extrémních hodnot je tzv. *Fisherův-Tippetův teorém* prezentovaný již v roce 1928, který byl následně v roce 1943 doplněn o *Gnědenkův teorém*. Hlavní myšlenkou tohoto teorému je, že pokud existuje posloupnost konstant  $\{a_n > 0\}$  a  $\{b_n\}$  takových, že

$$\Pr\left\{\frac{(M_n - b_n)}{a_n} \leq y\right\} \rightarrow G(y), \quad n \rightarrow \infty, \quad (3.35)$$

kde  $G$  představuje nedegenerovanou distribuční funkci, pak funkce  $G$  následuje jednu ze tří níže vymezených skupin rozdělení extrémních hodnot:

- typ I – *Gumbelovo rozdělení* definované vztahem

$$G(y) = \exp\left\{-\exp\left[-\left(\frac{y-b}{a}\right)\right]\right\}, \quad -\infty < y < \infty; \quad (3.36)$$

- typ II – *Fréchetovo rozdělení* v podobě

$$G(y) = \begin{cases} 0, & y \leq b, \\ \exp\left\{-\left(\frac{y-b}{a}\right)^{-\alpha}\right\}, & y > b; \end{cases} \quad (3.37)$$

- typ III – *Weibullovo rozdělení* stanovené pomocí rovnice

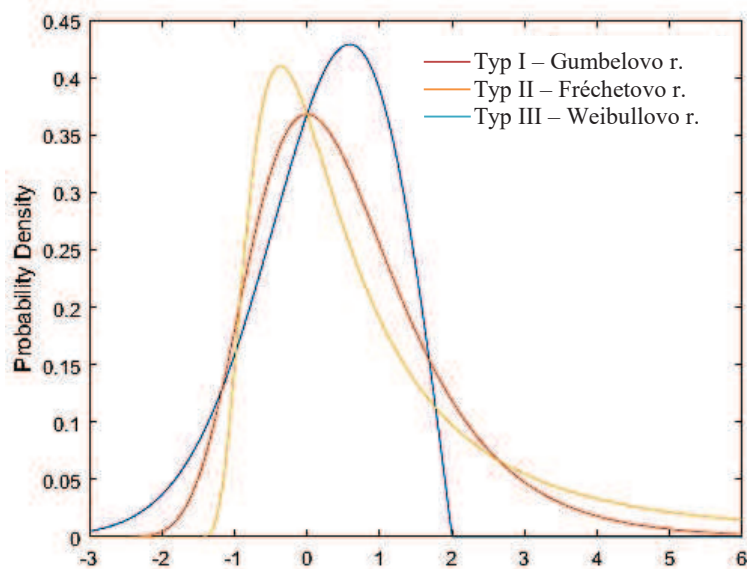
$$G(y) = \begin{cases} \exp\left\{-\left[\left(\frac{y-b}{a}\right)^\alpha\right]\right\}, & y < b, \\ 1, & y \geq b, \end{cases} \quad (3.38)$$

přičemž parametry  $a$  a  $\alpha$  musí nabývat hodnot větších než 0.

Z výše uvedeného můžeme vyvodit, že výběrová maxima z množiny  $M_n^*$  konvergují k rozdělení proměnných pocházejících z jedné ze skupin rozdělení extrémních hodnot. Každá skupina ze zmíněných tří typů rozdělení je charakteristická mírou polohy ( $b$ ) a mírou variability ( $a$ ), Fréchetovo a Weibullovo rozdělení je navíc popsáno parametrem  $\alpha$ . Pozoruhodným poznatkem této teorie je skutečnost, že Gumbelovo, Fréchetovo a Weibullovo rozdělení jsou jedinými možnými hranicemi pro rozdělení množiny  $M_n^*$ , a to bez ohledu na tvar populační distribuční funkce  $F$ . Porovnáme-li si principy tohoto teorému s centrální limitní větou, zjistíme jejich vzájemnou podobnost, viz Coles (2001).

Již dopředu můžeme předpokládat, že každá z výše popsaných skupin rozdělení se určitým způsobem liší ve svém průběhu, a to zejména v koncových oblastech. Tyto tři typy rozdělení jsou zachyceny pomocí Obr. 3.1 níže.

Obr. 3.1: Funkce hustoty pravděpodobnosti jednotlivých typů rozdělení



Zdroj: <http://www.mathworks.com/help/stats/examples/modelling-data-with-the-generalized-extreme-value-distribution.html>

Pokud se blíže zaměříme na chování limitního rozdělení funkce hustoty pravděpodobnosti v jejím pravém koncovém bodě  $y^*$ , zjistíme, že u Weibullova rozdělení je horní bod  $y^*$  konečným bodem, kdežto v případě Fréchetova a Gumbelova rozdělení platí rovnost bodu  $y^*$  nekonečnu. Další odlišností je rozložení funkce hustoty, které má exponenciální charakter pro rozdělení typu I a polynomický charakter pro rozdělení typu II. Abychom se v rámci modelování vyhnuli volbě odpovídajícího typu rozdělení extrémních hodnot na základě subjektivního úsudku, došlo ke sjednocení těchto tří skupin rozdělení vytvořením tzv. *zobecněného rozdělení extrémních hodnot* (GEV rozdělení) přiblíženého v následující kapitole.

### 3.5.1 ZOBECNĚNÉ ROZDĚLENÍ EXTRÉMNÍCH HODNOT

Sloučením Gumbelova, Fréchetova a Weibullova rozdělení získáme tedy zobecněné rozdělení extrémních hodnot, které je definováno vztahem

$$G(y) = \exp \left\{ - \left[ 1 + \xi \left( \frac{y - \nu}{\sigma} \right)^{-\frac{1}{\xi}} \right] \right\}, \quad 1 + \xi \left( \frac{y - \nu}{\sigma} \right) > 0, \quad (3.39)$$

kde jednotlivé parametry musí splňovat podmínky  $-\infty < \nu < \infty$ ,  $\sigma > 0$  a  $-\infty < \xi < \infty$ . Toto rozdělení je specifikováno třemi parametry – parametrem polohy  $\nu$ , parametrem variability  $\sigma$  a parametrem tvaru  $\xi$ . Právě prostřednictvím hodnoty parametru  $\xi$  je určováno chování konce neboli chvostu daného rozdělení. V případě, že  $\xi > 0$ , zobecněné rozdělení extrémních hodnot spadá do kategorie Fréchetova rozdělení. Pokud platí, že  $\xi < 0$ , tak *GEV* rozdělení následuje Weibullovo rozdělení. Je-li  $\xi = 0$ , tak pracujeme s rozdělením Gumbelova typu. Čím je hodnota odhadovaného parametru  $\xi$  vyšší, tím jsou chvosty rozdělení těžší.

### 3.5.2 PŘÍSTUP PEAK OVER THE THRESHOLD

V oblasti statistiky je k dispozici celá řada různých metod, které lze aplikovat pro modelování extrémních hodnot, přičemž často jsou v praxi vyžívány přístupy *metoda blokových maxim* a *metoda Peak over the threshold* (*POT* metoda).

Dle Coles (2001) je metoda blokových maxim založena na seskupení dat do bloků o stejné délce a následném nasazení zobecněného rozdělení extrémních hodnot na množinu blokových maxim. Kritickým bodem této metody je stanovení velikosti bloku, jelikož příliš malé bloky předznamenávají vyšší pravděpodobnost špatného odhadu modelu a naopak příliš velké bloky generují nízký počet blokových maxim, což rovněž může vést k významným odchylkám v odhadu. Nejlepším způsobem je proto určit délky bloků v podstatě přirozenou cestou, kdy je jejich velikost dána zvoleným časovým obdobím (např. měsíc, čtvrtletí či rok). Za hlavní nevýhodu této metody lze považovat fakt, že se zabývá analyzováním pouze maximálních hodnot datových bloků a nebere v potaz další vysoké hodnoty, které by se také daly považovat za extrémní.

Druhým zmíněným přístupem k modelování extrémních hodnot je metoda Peak over the threshold, jejímž výstupem jsou modely analyzující náhodné veličiny, které přesahují určitou prahovou hodnotu. *POT* metoda byla detailně rozpracována v roce 1990 pány Davidsonem a Smithem, avšak předtím byly její principy obsaženy již v práci pana Pickandse datující se k roku 1975. Tato metoda je velmi často aplikována na modelování pojistných, ale také finančních dat, a proto je východiskem pro tvorbu regresního modelu i v případě této diplomové práce.

Metoda Peak over the threshold je založena na předpokladu zobecněného Paretova rozdělení (*GP* rozdělení) náhodných veličin  $y_i$ . Pokud jsou modely dle *POT* přístupu správně zkonstruovány, zachycují extrémní hodnoty lépe než modely následující metodu blokových maxim. Klíčovým krokem *POT* metody je stanovení prahové hodnoty, neboť její špatné určení může vést k výrazným odchylkám ve výsledném odhadu. Pro zvolení prahové

proměnné mohou být využity různé techniky a nástroje, zejména tzv. mean residual life plot či kvantilová regrese, se kterou pracujeme v praktické části této práce.

Samotný odhad parametrů modelu lze pak v případě přístupu Peak over the threshold provést například prostřednictvím metody maximální věrohodnosti, metody vážených momentů nebo metody percentilů, blíže viz Beirlant, Goegebeur, Segers, Teugels (2004). V diplomové práci je odhad parametrů zpracován pomocí metody maximální věrohodnosti, stejně jako u regresního gamma modelu.

### 3.6 REGRESNÍ ZOBECNĚNÝ PARETŮV MODEL

Jak jsme již dříve zmínili, pro modelování individuálních pojistných škod lze také uplatnit zobecněný lineární model, který pracuje s předpokladem smíšeného rozdělení pravděpodobnosti. Hodnoty  $y_i$  nepřesahující prahovou proměnnou jsou modelovány pomocí dříve popsaneho regresního log-gamma modelu, zatímco hodnoty nad stanovenou prahovou proměnnou následují zobecněné Paretovo rozdělení, a opírají se tedy o východiska teorie extrémních hodnot, zejména metody Peak over the threshold. V této části se zaměřujeme pouze na konstrukci regresního zobecněného Paretova modelu. Nejdříve je důležité charakterizovat *GP* rozdělení, na které následně nasadíme daný regresní Paretův model.

#### 3.6.1 ZOBECNĚNÉ PARETOVO ROZDĚLENÍ

Uvažujeme-li náhodnou veličinu  $y_i$ , jejíž rozdělení odpovídá zobecněnému Paretovu rozdělení, pak kumulativní distribuční funkci tohoto rozdělení můžeme zachytit pomocí rovnice

$$G(y; \xi, \nu, \sigma) = \begin{cases} 1 - \left(1 + \xi \cdot \frac{y - \nu}{\sigma}\right)^{-\frac{1}{\xi}}, & 1 + \xi \cdot \frac{y - \nu}{\sigma} > 0, \xi \neq 0, \\ 1 - \exp\left(-\frac{y - \nu}{\sigma}\right), & y - \nu > 0, \xi = 0, \end{cases} \quad (3.40)$$

kde symbol  $\xi$  značí parametr tvaru,  $\nu$  je parametr polohy a  $\sigma$  představuje parametr variability, přičemž prahová proměnná tihne k pravému koncovému bodu této distribuční funkce.

Funkci hustoty pravděpodobnosti  $g$  lze pro zobecněné Paretovo rozdělení zapsat prostřednictvím vztahu

$$g(y; \xi, \nu, \sigma) = \begin{cases} \frac{1}{\sigma} \cdot \left(1 + \xi \cdot \frac{y - \nu}{\sigma}\right)^{-\frac{1}{\xi} - 1}, & 1 + \xi \cdot \frac{y - \nu}{\sigma} > 0, \xi \neq 0, \\ \frac{1}{\sigma} \cdot \exp\left(-\frac{y - \nu}{\sigma}\right), & y - \nu > 0, \xi = 0. \end{cases} \quad (3.41)$$

Mezi zobecněným Paretovým rozdělením a *GEV* rozdělením můžeme shledat určité podobnosti, hlavně co se týká parametru  $\xi$ . I v případě *GP* rozdělení je totiž parametr tvaru dominantním faktorem pro zjištění jeho kvalitativního chování v koncové oblasti, stejně jako u *GEV* rozdělení, blíže viz Coles (2001).

### 3.6.2 ODVOZENÍ ZOBECNĚNÉHO PARETOVA MODELU

Pokud parametry zobecněného Paretova rozdělení vyjádříme prostřednictvím funkcí individuálních regresních faktorů, získáme regresní model vhodný pro modelování pojistných škod. Máme-li tedy k dispozici nezávislou náhodnou proměnnou  $y_1, \dots, y_n$  a vektor rizikových faktorů  $x_i$ , lze vztah mezi vysvětlovanou proměnnou  $y_i$  a příslušnými vysvětlujícími proměnnými  $x_i$  zapsat způsobem

$$y_i | x_i \sim GP(\sigma(x_i), \xi(x_i)), \quad i = 1, \dots, n. \quad (3.42)$$

Funkce rizikových faktorů  $\sigma(x_i)$  a  $\xi(x_i)$  můžeme pak stanovit jako kombinace

$$\sigma = z_1(x_i, \alpha_1) \quad (3.43)$$

a

$$\xi = z_2(x_i, \alpha_2), \quad (3.44)$$

kde  $\alpha = (\alpha_1, \alpha_2)$  představuje vektor regresních koeficientů pro regresní *GP* model.

### 3.6.3 ODHAD PARAMETRŮ ZOBECNĚNÉHO PARETOVA MODELU

Odhady parametrů *GP* modelu můžeme získat obdobným postupem jako v případě regresního gamma modelu. Opět využijeme metodu maximální věrohodnosti, jejíž principy jsou přiblíženy již v kapitole 3.2. Vycházíme-li z teoretických poznatků této metody, můžeme věrohodnostní funkci po logaritmické transformaci pro regresní *GP* model zaznamenat v podobě

$$l(y; \xi, \nu, \sigma) = \sum_{i=1}^n \log g(y_i; \xi(x_i), \nu, \sigma(x_i)). \quad (3.45)$$

Pokud uvažujeme funkci hustoty  $g$  definovanou pomocí vztahu (3.41), pak zlogaritmovaná věrohodnostní funkce  $l$  může být přepsána do tvaru

$$l(y; \xi, \nu, \sigma) = \sum_{i=1}^n -\log \sigma(x_i) - \left( \frac{1}{\xi(x_i)} + 1 \right) \cdot \log \left( 1 + \xi(x_i) \cdot \frac{y_i - \nu}{\sigma(x_i)} \right). \quad (3.46)$$

V regresním modelu za předpokladu *GP* rozdělení je rovněž nezbytné zachytit příslušnou prahovou proměnnou, neboť je pro modelování extrémních hodnot aplikována *POT* metoda. Jelikož je výše pojistné škody obvykle odvislá od celkové hodnoty



pojišťovaného majetku, je vhodné vyjádřit prahovou proměnnou jako funkci rizikových faktorů, díky čemuž se bude její hodnota měnit v rámci individuálních charakteristik konkrétních pojistných smluv. V regresním *GP* modelu je prahová proměnná obsažena v parametru polohy  $\nu$ , který můžeme rozšířit uplatněním funkční závislosti určené jako

$$\nu = z_3(x_i, \gamma), \quad (3.47)$$

kde symbol  $\gamma$  představuje odhadovaný koeficient funkce pro prahovou proměnnou. V návaznosti na funkční vyjádření prahové proměnné je v diplomové práci odhadován vektor koeficientů v podobě  $\alpha = (\alpha_1, \alpha_2, \gamma)$  pro případ zobecněného Paretova modelu.

Jak tvrdí Beirlant, Goegebeur, Segers, Teugels (2004), odhad příslušného tvaru prahové proměnné, konkrétně koeficientu  $\gamma$ , můžeme jednoduše vyřešit prostřednictvím kvantilové regrese. Důležité je předem určit procento  $q$  kvantilové regrese, přičemž musí platit podmínka  $0 < q < 1$ . Tuto optimalizační úlohu lze zapsat pomocí rovnice

$$\min_{\gamma_p} \sum_{i=1}^n \left( q(y_i - \nu(x_i))^+ + (1-q) \cdot (y_i - \nu(x_i))^- \right), \quad (3.48)$$

kde  $x^+ = \max(0, x)$  a  $x^- = \max(0, -x)$ . Odhad koeficientu  $\gamma$  je společně s funkčním předpisem prahové proměnné  $\nu(x_i)$  dosazen do rovnice (3.46), díky čemuž následně získáme odhad regresních koeficientů  $\alpha_1$  a  $\alpha_2$ .

Pro posouzení kvality a přesnosti vytvořeného zobecněného Paretova modelu se v matematické praxi obvykle konstruuje exponenciální kvantilový graf. Podkladovými veličinami pro vytvoření tohoto grafu jsou hodnoty proměnné  $R_i$  vypočtené dle matematického vzorce

$$R_i = \begin{cases} \frac{1}{\xi(x_i)} \cdot \log \left( 1 + \xi(x_i) \cdot \frac{y_i}{\sigma(x_i)} \right), & \xi(x_i) \neq 0, \\ \frac{y_i}{\sigma(x_i)}, & \xi(x_i) = 0, \end{cases} \quad (3.49)$$

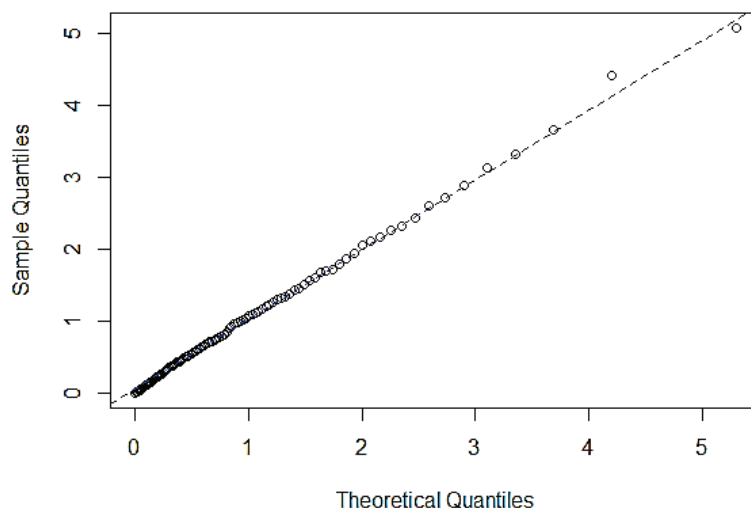
přičemž z hodnot  $R_i$  jsou následně odvozeny souřadnice exponenciálního kvantilového grafu prostřednictvím vztahu

$$\left( -\log \left( 1 - \frac{i}{n+1} \right), R_i \right), \quad i = 1, \dots, n, \quad (3.50)$$

kde proměnnou  $R_i$  je před sestrojením grafu nezbytné seřadit způsobem  $R_1 \leq R_2 \leq \dots \leq R_n$ , blíže viz Coles (2001).

Výsledný exponenciální kvantilový graf nazývaný také jako tzv.  $Q-Q$  plot by se měl blížit podobě zachycené pomocí ilustračního Obr. 3.2 níže, neboť v tomto případě lze hovořit o správně odhadnutém a poměrně přesném regresním modelu odpovídajícím zkoumaným datům.

Obr. 3.2: Exponenciální kvantilový graf



Zdroj: <http://stackoverflow.com/questions/23196088/how-can-i-generate-an-exponential-q-q-plot-in-r>

V případě, že regresní  $GP$  model je přesný a dobře odpovídá pozorovaným datům, pak se jednotlivé body grafu přibližují pomocné přímce vedoucí pod úhlem 45 stupňů z výchozího nulového bodu, a to podobně jako ve výše zobrazeném vzorovém grafu. Naopak čím více jsou znázorněné body vzdálené od dané jednotkové přímky, tím méně je regresní model vhodný pro modelování analyzovaných dat.

### 3.7 REGRESNÍ SMÍŠENÝ MODEL

Pro účely praktické části je zde vhodné nastínit princip smíšeného modelu tak, jak je uvažován v této diplomové práci. Jelikož aplikací metody Peak over the threshold dochází v podstatě k rozdělení datového vzorku vysvětlované proměnné na dva soubory, kdy jeden je modelován regresním log-gamma modelem a druhý zobecněným Paretovým modelem, je nezbytné tyto dva modely určitým způsobem sloučit v jeden celek. Problematika smíšeného rozdělení je poměrně obtížná, a proto využijeme základního vztahu pro spojení dvou rozdílných rozdělení pravděpodobnosti, a tedy i modelů, do jednoho smíšeného rozdělení, a to prostřednictvím zápisu

$$f(y) = p \cdot f(\Phi_1) + (1 - p) \cdot f(\Phi_2), \quad (3.51)$$

kde symbolem  $p$  je myšlena pravděpodobnost, že daná vysvětlovaná proměnná  $y_i$  bude následovat rozdělení pravděpodobnosti  $\Phi_1$ , v opačném případě pak rozdělení pravděpodobnosti  $\Phi_2$ . Tento vztah lze také uplatnit pro analýzu střední hodnoty náhodné veličiny  $y_i$ , kdy jej můžeme upravit do tvaru

$$E[y] = p \cdot E[y_1] + (1 - p) \cdot E[y_2], \quad (3.52)$$

kde  $y_1$  a  $y_2$  představují jednotlivé vektory náhodných hodnot po rozdělení dle prahové proměnné. Jelikož jsou v této práci modelovány individuální pojistné škody, přepíšeme rovnici výše do tvaru

$$E[y] = p \cdot E[LC] + (1 - p) \cdot E[SC], \quad (3.53)$$

kde pojmem  $LC$  označíme extrémní ztráty a symbolem  $SC$  ztráty malé. Pojistné škody převyšující prahovou proměnnou ( $LC$ ) následně modelujeme dle zobecněného Paretova modelu, na individuální pojistné škody nepřesahující prahovou proměnnou ( $SC$ ) pak aplikujeme regresní log-gamma model. Parametr  $E[SC]$  zjistíme po provedení optimalizace z rovnice (3.31) dosazením odhadnutých koeficientů  $\beta$ , parametr střední hodnoty pro zobecněné Paretovo rozdělení pak určíme pomocí vzorce

$$E[LC] = \nu + \frac{\sigma}{1 - \xi}. \quad (3.54)$$

Výše popsané principy jsou využity zejména při porovnání a zhodnocení kvality a přesnosti jednotlivých regresních modelů.

### 3.8 REZIDUÁLNÍ ANALÝZA

Velmi důležitým krokem v rámci finančního modelování je ověření přesnosti a významnosti daného regresního modelu. Pro posouzení kvality zobecněného lineárního modelu lze použít různé techniky a statistiky, přičemž nejčastěji se provádí reziduální analýza doplňovaná o výpočet zvolených kritériálních statistik. V této diplomové práci je přesnost modelu testována pomocí analýzy reziduí, konkrétně se zaměříme na hrubá rezidua a Pearsonova rezidua, jelikož jsou vhodným nástrojem pro skupinu  $GLM$ , jejich výpočet je poměrně snadný a lze je také dále modifikovat – např. můžeme provést jejich standardizaci či studentizaci. Vypočtená rezidua nám pak poskytnou informace o naměřených odchylkách mezi jednotlivými pozorovanými a odhadovanými hodnotami.

*Hrubá rezidua* můžeme velmi jednoduše vypočítat dle vzorce

$$r_i^R = y_i - \hat{\mu}_i, \quad (3.55)$$

kde  $y_i$  jsou skutečné pozorované hodnoty vysvětlované proměnné a  $\hat{\mu}_i$  je označením pro jednotlivé očekávané střední hodnoty. Jedná se tedy v podstatě o rozdíl skutečných hodnot a odhadnutých hodnot.

*Pearsonova rezidua* lze pak získat ze vztahu

$$r_i^P = \frac{y_i - \hat{\mu}_i}{\sqrt{V(\hat{\mu}_i)}}, \quad (3.56)$$

kde  $y_i$  představuje skutečné hodnoty závislé veličiny,  $\hat{\mu}_i$  jednotlivé predikované střední hodnoty a  $V(\hat{\mu}_i)$  příslušný rozptyl. Pokud jsou výsledkem vysoké hodnoty  $r_i^P$  v absolutním vyjádření, máme špatně nastavený regresní model neodpovídající dostatečně skutečným pozorováním.

V praktické části diplomové práce je přesnost a kvalita odhadnutých regresních modelů, konkrétně log-gamma modelu a smíšeného modelu, porovnána prostřednictvím grafů zachycujících vyhlazená rezidua společně s příslušnými hranicemi konfidenčních intervalů za předpokladu 95% hladiny spolehlivosti.

## 4 ODHAD A VYHODNOCENÍ REGRESNÍHO MODELU

Záměrem praktické části práce je vytvořit a odhadnout vhodný regresní model analyzující vztah mezi velikostí pojistné škody související s jednou smlouvou a uvažovanými rizikovými faktory. V diplomové práci je tento problém řešen prostřednictvím dvou přístupů ke konstrukci regresního modelu, přičemž výsledky aplikace jsou postupně obsaženy v této části práce.

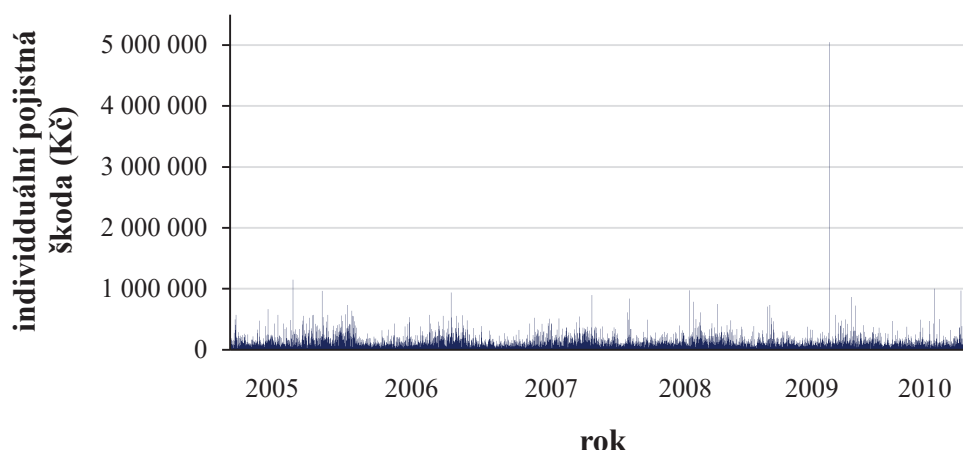
V úvodních odstavcích této kapitoly je nezbytné přesně vymezit vstupní data, která slouží jako podkladový materiál pro modelování individuálních pojistných škod. Poté se již zabýváme samotným odhadem regresních modelů. Nejdříve se soustředíme na vytvoření zobecněného lineárního modelu následujícího gamma rozdělení pravděpodobnosti a následně se zaměřujeme na konstrukci zobecněného lineárního modelu za předpokladu smíšeného rozdělení pravděpodobnosti, které zahrnuje *GP* rozdělení a gamma rozdělení. Důležitou fází v procesu tvorby druhého zmiňovaného modelu je stanovení úrovně prahové proměnné, která je v práci zjištěna prostřednictvím kvantilové regrese. Na základě odvození prahové proměnné je pak rozdělen vektor vysvětlované veličiny (*severity*) na dva vektory, kdy v jednom vektoru jsou zachyceny pojistné škody do prahové proměnné a ve druhém vektoru jsou obsaženy škody nad prahovou proměnnou. Hodnoty těchto dvou vektorů jsou pak modelovány rozdílně. Na závěr jsou výsledné odhady porovnány různými technikami. Zohledňován je jednak vývoj Pearsonových reziduí u uvažovaných modelů, ale analyzována jsou také hrubá rezidua a směrodatné odchylky. Na základě těchto kritérií jsou pak oba modely celkově zhodnoceny a je posouzena jejich kvalita, přesnost a vhodnost použití pro modelování vybraných dat.

K jednotlivým výpočtům optimalizačních úloh je použit program MATLAB verze R2010b, k doplňkovým propočtům pak také program STATA/SE verze 11.0. Výsledné grafy a tabulky jsou zpracovány převážně v programu MS Excel.

### 4.1 ANALÝZA VSTUPNÍCH DAT

Při odhadu regresního modelu vycházíme ze souboru dat, ve kterém jsou obsaženy reálné údaje z pojistných smluv týkajících se havarijního pojištění osobních vozidel. Předmětem regresní analýzy jsou individuální pojistné škody (*severity*), které vznikly v průběhu časového období 2005 – 2010. Celkově tedy pracujeme se souborem čítajícím 22 274 pozorování. V Graf 4.1 jsou znázorněny jednotlivé výše pojistných škod za sledované období.

Graf 4.1: Výše individuálních pojistných škod v období 2005 - 2010



Již nyní můžeme z výše uvedeného grafu usuzovat, že některé hodnoty pojistných škod představovaly pro danou pojišťovnu výraznější ztrátu, konkrétně ve 47 případech výše pojistné škody překročila hranici 500 000 Kč, ve 2 případech pak hranici 1 000 000 Kč, a to z celkových 22 274 pozorování. Dále je patrné, že v roce 2009 vznikla dokonce ojedinělá pojistná škoda ve výši téměř 5,1 mil. Kč. Podíváme-li se blíže na zkoumaná data, zjistíme, že tuto extrémní hodnotu lze odůvodnit poměrně vysokou cenou pojištěného vozidla, která činila cca 7,5 mil. Kč.

V rámci modelování pojistných škod jsou uvažovány vysvětlující proměnné jak ve formě spojité veličiny, tak rovněž v podobě kategorické veličiny. V případě spojitých veličin pracujeme s těmito rizikovými faktory:

- stáří vozidla (*agecar*),
- věk majitele vozidla (*ageman*),
- cena vozidla (*carprice*),
- síla motoru (*kwvol*) kombinující objem motoru (*vol*) a výkon motoru (*kw*) prostřednictvím vztahu  $kw/vol \cdot 1000$ ,
- počet obyvatel města (*nocit*).

Do modelu jsou dále zahrnuty tyto regresní faktory, které jsou zapsané v kategorické formě:

- pohlaví pojistníka (*gender*) – muž/žena,
- forma vlastnictví vozidla (*company*) – osobní/slужеbní,
- typ paliva (*fuel*) – benzín/nafta/LPG/ostatní.

V Tab. 4.1 jsou zaznamenány vypočtené základní statistické znaky jednotlivých spojitých veličin, včetně charakteristik vysvětlované proměnné (*severity*).

Tab. 4.1: Základní statistické charakteristiky pozorovaných dat

Rizikový faktor	Střední hodnota	Směrodatná odchylka	Šikmost	Špičatost
<i>severity</i>	52 210,12	72 625,59	17,50	1026,25
<i>agecar</i>	4,28	2,93	0,78	4,31
<i>ageman</i>	46,02	13,94	0,33	2,41
<i>carprice</i>	474 959,11	407 200,16	4,99	53,92
<i>kwvol</i>	44,25	8,95	0,70	5,44
<i>nocit</i>	608 348,31	539 635,81	0,32	1,15

Z této tabulky je zřejmé, že všechny zkoumané spojité veličiny mají nenormální rozdělení. Koeficienty šikmosti jsou kladné, což svědčí o pravostranně orientovaném rozdělení. Z hodnot koeficientů špičatosti lze vyvodit, že rozdělení je špičatější než normální rozdělení. Těmito výsledky je tudíž potvrzen předpoklad, že pojistné škody by měly být modelovány na pravostranném zešikmeném spojitým rozdělení, což je zvoleným log-gamma i Paretoovým rozdělením splněno.

Pomocí aplikace přístupu *GLM* můžeme tedy v dalších kapitolách vytvořit a odhadnout regresní model za předpokladu log-gamma rozdělení a smíšeného rozdělení. Díky této metodě je možné vyjádřit funkční závislost velikosti individuální pojistné škody na zmíněných regresních faktorech, ať už v podobě kategorických či spojitých veličin.

## 4.2 KONSTRUKCE REGRESNÍHO LOG-GAMMA MODELU

Dříve než přistoupíme k samotnému odhadu parametrů regresního modelu, je potřeba určitým způsobem vybrat ze souboru dat ty rizikové faktory, které ovlivňují velikost individuální pojistné škody a jsou statisticky významné. K vyřešení tohoto problému použijeme výsledky jednorozměrné analýzy, které následně podrobíme Z-testu. Na základě zjištěných hodnot poté zahrneme do modelu pouze ty faktory, jež jsou statisticky významné, a provedeme konečný odhad parametrů celkového modelu pomocí metody maximální věrohodnosti. Pro zjednodušení používáme v dalších částech textu pro regresní log-gamma model také označení symbolem  $M_1$ .

### 4.2.1 JEDNOROZMĚRNÁ ANALÝZA NA BÁZI Z-TESTU

Z datového vzorku máme k dispozici osm uvažovaných regresních faktorů, které musíme podrobit zjednodušující jednorozměrné analýze dat a následné statistické verifikaci jejich významnosti.

Prvním krokem je provedení odhadu parametrů regresního modelu metodou maximální věrohodnosti, a to pro každý rizikový faktor zvlášť dle příslušného vztahu (3.29), kdy do log-link funkce definované vzorcem (3.31) dosadíme za vektor  $X$  hodnoty pouze jednoho zkoumaného rizikového faktoru. Po spuštění optimalizační úlohy v prostředí programu MATLAB získáme odhad parametrů  $\beta$  (konkrétně konstanty  $\beta_0$  a regresního koeficientu  $\beta_1$ ) a parametru disperze  $\phi$  společně s odpovídajícími chybami odhadu  $SE$ , které zjistíme z informační matice vzniklé převedením matice  $H$  vypočtené dle rovnice (3.9). Tento postup je aplikován opakovaně na všech osm analyzovaných regresních faktorů. Výsledky této analýzy dat jsou zachyceny v Příloze č. 1.

Na základě výsledných odhadů dosažených jednorozměrnou analýzou je dále ověřována statistická významnost jednotlivých rizikových faktorů, a to prostřednictvím Z-testu. Pro každou z osmi provedených jednorozměrných regresí je nezbytné stanovit odpovídající Z-statistiku, příslušné hranice konfidenčních intervalů a  $p$ -hodnotu, a to dle vztahů (3.12) a (3.13). Pomocí Tab. 4.2 níže jsou zaznamenány  $p$ -hodnoty získané využitím funkcí integrovaných v programu MS Excel za předpokladu 5% hladiny významnosti, na základě kterých je poté vyhodnocena statistická významnost jednotlivých regresních faktorů a jejich zahrnutí do celkového modelu  $M_1$ .

Tab. 4.2: Vypočtené  $p$ -hodnoty ověřující významnost faktorů

Rizikový faktor		$p$ -hodnota koeficientu $\beta_1$
stáří vozidla ( <i>agecar</i> )		0,0000
věk majitele vozu ( <i>ageman</i> )		0,0003
cena vozidla ( <i>carprice</i> )		0,0000
forma vlastnictví vozu ( <i>company</i> )		0,0000
pohlaví pojistníka ( <i>gender</i> )		0,0000
síla motoru ( <i>kwvol</i> )		0,0168
počet obyvatel města ( <i>nocit</i> )		0,0000
typ paliva ( <i>fuel</i> )	1. kategorie	0,0000
	2. kategorie	0,0000
	3. kategorie	0,0000

Z výsledků můžeme usuzovat, že všech osm uvažovaných rizikových faktorů určitým způsobem ovlivňuje velikost individuálních pojistných škod, neboť veškeré  $p$ -hodnoty jsou nižší než hladina významnosti nastavená na úrovni 0,05. Do celkového regresního log-gamma



modelu je tedy vhodné zahrnout všechny tyto rizikové faktory a považovat je za vysvětlující proměnné. Toto tvrzení je navíc podpořeno dosaženými hodnotami  $Z$ -statistiky a hranicemi odpovídajících konfidenčních intervalů, které jsou obsaženy v Příloze č. 2.

#### 4.2.2 ODHAD PARAMETRŮ MODELU

Dle výsledků dosažených v předešlé kapitole tedy do celkového modelu zahrneme vektor hodnot všech rizikových faktorů  $x_i$ . Log-link funkci vyjadřující parametr střední hodnoty vytvoříme na základě vztahu (3.31), výsledné odhady pak získáme optimalizací úlohy zachycené vzorcem (3.29), respektive (3.32). Vypočtené odhady regresních parametrů  $\mu$  a  $\phi$  jsou zobrazeny v Tab. 4.3 níže. Parametr střední hodnoty je zde zachycen pomocí kombinace odhadnutých regresních koeficientů  $\beta_i$  a odpovídajících vysvětlujících proměnných, a to na základě zvolené log-link funkce. Míra variability nabývá konstantní hodnoty pro veškerá pozorování.

Tab. 4.3: Výsledný odhad parametrů regresního log-gamma modelu  $M_1$

Rizikový faktor		Koeficient	Střední hodnota $\mu$	Parametr disperze $\phi$
-		konstanta $\beta_0$	10,5136	0,8957
agecar		$\beta_1$	-0,0235	
ageman		$\beta_2$	-0,0079	
carprice		$\beta_3$	$3,4408 \cdot 10^{-7}$	
company		$\beta_4$	-0,5800	
gender		$\beta_5$	0,0878	
kwvol		$\beta_6$	0,0109	
nocit		$\beta_7$	$4,0418 \cdot 10^{-8}$	
fuel	1. kategorie	$\beta_8$	0,3045	
	2. kategorie	$\beta_9$	-0,6118	
	3. kategorie	$\beta_{10}$	0,6169	

Z tabulky výše je patrné, že při zahrnutí takového počtu vysvětlujících proměnných do modelu je vliv jednotlivých dílčích faktorů na celkovou výši individuální pojistné škody poměrně malý. Provést porovnání jednotlivých vazeb a vztahů mezi vysvětlovanou proměnnou a regresními faktory je poměrně složité, neboť každý faktor je uváděn v jiných jednotkách či formě zápisu, a tudíž je obtížné z těchto hodnot vyvodit nějaké konkrétní závěry o síle závislosti. Na první pohled silnější vazby lze pozorovat u proměnných *fuel* a *company*,

nicméně se jedná o kategorické proměnné, tudíž vektor těchto faktorů je tvořen pouze 0 nebo 1. Naopak v případě kategorií *carprice* a *nocit* by se mohlo zdát, že se jedná o velmi nízké bezvýznamné hodnoty koeficientů, avšak to je způsobeno zejména tím, že v rámci těchto dvou proměnných pracujeme s čísly v řádu desítek až stovek tisíc. Dle hodnot jednotlivých regresních koeficientů lze sledovat pozitivní i negativní směr závislosti vůči vysvětlované proměnné (*severity*). Parametr disperze je konstantní na úrovni 0,8957, tudíž je zde splněna podmínka kladné hodnoty parametrů gamma rozdělení, a to jak v případě parametru  $\phi$ , tak také v případě parametru  $\mu$ , který nemůže být záporný z důvodu použití log-link funkce.

Přesnost a kvalita modelu  $M_1$  je ověřena hlavně prostřednictvím analýzy hrubých reziduí a Pearsonových reziduí, která je společně s dalším zhodnocením modelu obsažena v kapitole 4.4.

### 4.3 KONSTRUKCE REGRESNÍHO SMÍŠENÉHO MODELU

Jak již bylo v této práci několikrát zmíněno, k modelování pojistných škod lze přistoupit také z hlediska zohlednění extrémních ztrát. Pokud aplikujeme tuto metodiku, je nevyhnutelné pracovat se smíšeným rozdělením pravděpodobnosti za předpokladu gamma rozdělení a zobecněného Paretova rozdělení. Tento model tedy v diplomové práci pojmenujeme jako smíšený model, jenž se skládá z log-gamma modelu a zobecněného Paretova modelu, a označíme jej symbolem  $M_2$ .

V rámci smíšeného modelu pracujeme se stejnými regresními faktory jako v případě modelu  $M_1$ , a tudíž znovu neprovádíme jednorozměrnou analýzu a následnou statistickou verifikaci na bázi Z-testu. Předpokládáme tedy, že rizikové faktory jsou opět statisticky významné a měly by být zahrnuty do regresního modelu  $M_2$ . Jediným rozdílem oproti modelu  $M_1$  je skutečnost, že v případě kategorické proměnné typ paliva (*fuel*) již bereme v úvahu pouze 1. kategorii, neboť smíšený model je i tak velmi složitý a jedná se tedy pouze o zjednodušující krok, který by neměl mít žádný výrazný vliv na výslednou kvalitu a přesnost modelu.

Prvním krokem v rámci modelování individuálních pojistných škod dle teorie extrémních hodnot, tedy přístupu Peak over the threshold, je určení prahové proměnné. Prahovou proměnnou lze stanovit buď jako hodnotu konstantní pro celý model, nebo prostřednictvím funkční závislosti na jiných veličinách. Pro účely této práce odvodíme prahovou proměnnou z ceny vozidla, neboť se to jeví jako nejlepší možná varianta z nabízených rizikových faktorů. Jelikož je cena vozidla v případě jednotlivých pojistných smluv poměrně dost odlišná, vyjádříme prahovou proměnnou jako funkci ceny vozidla.

V případě smíšeného modelu je tedy předpokladem, že individuální prahové proměnné získáme jako procento z ceny vozidla na základě vztahu (3.47), tedy  $v = \gamma \cdot \text{carprice}$ .

Hodnota procentního koeficientu  $\gamma$  byla zjištěna kvantilovou regresí dle vzorce (3.48), kdy jsme uvažovali úroveň  $q$  ve výši 0,99. Po vyřešení tohoto optimalizačního problému v programu MATLAB jsme získali koeficient  $\gamma$  na úrovni 0,6424. Na základě zjištěných hodnot prahových proměnných jsme schopni rozdělit hodnoty vektoru individuálních pojistných škod (*severity*) na vektor obsahující pojistné škody do dané prahové proměnné (*LGMseverity*) a na vektor zachycující škody nad prahovou proměnnou (*GPMseverity*). Pojistné škody nepřesahující prahovou proměnnou modelujeme pomocí stávajícího regresního log-gamma modelu, pro individuální pojistné škody převyšující danou prahovou proměnnou je pak uplatněn zobecněný Paretův model.

Přesnost a kvalita smíšeného modelu  $M_2$  je vyhodnocena v kapitole 4.4, kde je také tento model porovnán s regresním modelem  $M_1$ .

#### 4.3.1 ODHAD PARAMETRŮ LOG-GAMMA MODELU

Princip řešení této úlohy je v podstatě stejný jako v případě odhadu parametrů modelu  $M_1$ , pouze se nám změnila vysvětlovaná proměnná, kdy nyní bereme v úvahu hodnoty nepřesahující prahovou proměnnou, tedy *LGMseverity*. Výsledky odhadnutých parametrů modelu jsou pak zachyceny v Tab. 4.4 níže.

Tab. 4.4: Odhad parametrů log-gamma modelu v rámci modelu  $M_2$

Rizikový faktor	Koeficient	Střední hodnota $\mu$	Parametr disperze $\phi$
-	konstanta $\beta_0$	10,5142	0,8883
<i>agecar</i>	$\beta_1$	-0,0232	
<i>ageman</i>	$\beta_2$	-0,0079	
<i>carprice</i>	$\beta_3$	$3,5594 \cdot 10^{-7}$	
<i>company</i>	$\beta_4$	-0,5785	
<i>gender</i>	$\beta_5$	0,0872	
<i>kwvol</i>	$\beta_6$	0,0106	
<i>nocit</i>	$\beta_7$	$4,2158 \cdot 10^{-8}$	
<i>fuel</i> (1. kategorie)	$\beta_8$	0,3022	

Z tabulky výše je zřejmé, že odhadnuté regresní koeficienty se spolu s variačním parametrem změnilo jen minimálně oproti odhadu modelu  $M_1$ . Jednak tím je doložen

předpoklad, že nezahrnutí 2. a 3. kategorie proměnné *fuel* nemá výrazný vliv na odhad parametrů modelu, a navíc lze z těchto výsledků usuzovat, že většina vzniklých individuálních pojistných škod nepřesahovala danou prahovou proměnnou, a tudíž modelovaná závislá proměnná se změnila jen v případě několika málo pojistných smluv, konkrétně u 400 smluv z celkových 22 274.

#### 4.3.2 ODHAD PARAMETRŮ ZOBECNĚNÉHO PARETOVA MODELU

Pokud předpokládáme, že extrémní hodnoty, tedy ty, které překračují individuální prahovou proměnnou, následují zobecněné Paretovo rozdělení, pak je k jejich modelování vhodné využít regresní GP model popsany v kapitole 3.6. V případě GP rozdělení nicméně pracujeme se třemi parametry – parametrem polohy  $\nu$ , tvaru  $\zeta$  a variability  $\sigma$ . Každý z těchto tří parametrů je navíc vyjádřen prostřednictvím link funkcí regresních faktorů, kdy parametr polohy je stanoven jako funkce ceny vozidla a odpovídá prahové proměnné, zbylé dva parametry jsou pak přepsány prostřednictvím log-link funkcí dle vzorců (3.43) a (3.44). Parametry tohoto modelu pak odhadneme na základě optimalizace vztahu (3.46). Řešení této úlohy je zobrazené v Tab. 4.5, konkrétně hodnoty regresních koeficientů  $\alpha_i$  pro příslušný rizikový faktor a parametr zobecněného Paretova rozdělení.

Tab. 4.5: Odhad parametrů zobecněného Paretova modelu v rámci modelu  $M_2$

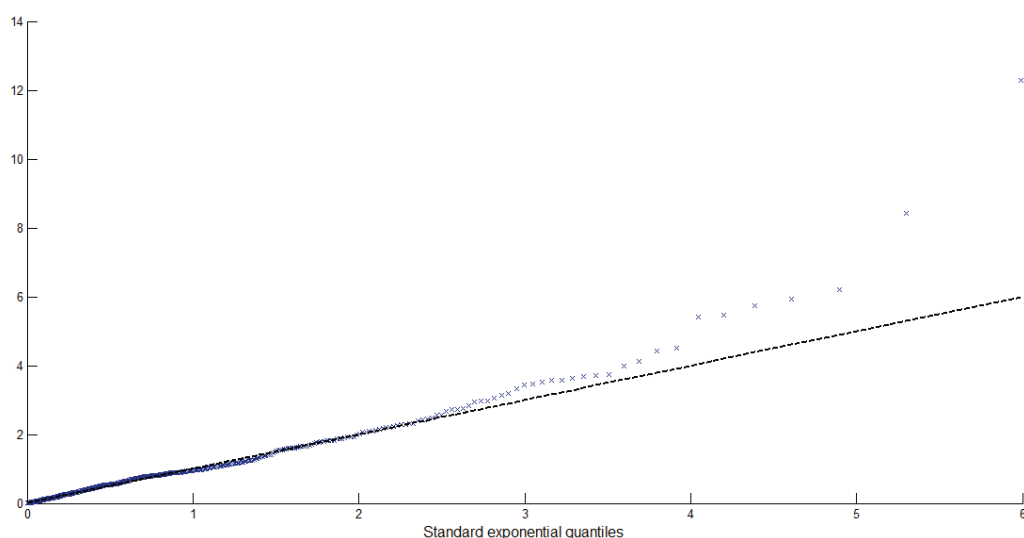
Rizikový faktor	Parametr tvaru $\zeta$	Parametr variability $\sigma$
- (konstanta)	-9,9850	8,7561
<i>agecar</i>	-0,0186	-0,1503
<i>ageman</i>	-0,1962	-0,2501
<i>carprice</i>	$1,2967 \cdot 10^{-7}$	$2,2764 \cdot 10^{-9}$
<i>company</i>	3,3951	-0,0685
<i>gender</i>	0,0124	-0,1091
<i>kwvol</i>	-0,0191	0,0283
<i>nocit</i>	$1,0314 \cdot 10^{-6}$	$1,7893 \cdot 10^{-7}$
<i>fuel</i> (1. kategorie)	-0,1189	0,2384

Parametry GP modelu byly odhadovány v závislosti na vysvětlované proměnné *GPMseverity*, která zachycuje pojistné škody přesahující individuální prahovou proměnnou, což představuje vzorek dat čítající pouze 400 pozorování. Vypočtené odhady koeficientů jsou opět poměrně nízká čísla vyjadřující jak pozitivní, tak negativní závislost. Abychom získali co nejlepší odhady koeficientů, bylo nezbytné věnovat hodně času nastavení výchozího řešení.

Odhad parametrů byl proveden postupným zahrnováním jednotlivých rizikových faktorů do modelu, díky čemuž jsme získali alespoň orientační hodnotu odhadovaného regresního koeficientu, kterou jsme následně mohli použít ve výchozím řešení. Snahou bylo vyhnout se lokálnímu minimu, a tudíž jsme byli nuceni spouštět optimalizační úlohu opakovaně na základě změněného výchozího řešení a hledat co nejlepší odhad modelu.

Abychom byli schopni vyhodnotit kvalitu a přesnost odhadu zobecněného Paretova modelu, vytvořili jsme v prostředí MATLAB exponenciální kvantilový graf, a to dle postupu následujícího vztahu (3.49) a (3.50), který je zobrazen v Graf 4.2 níže.

*Graf 4.2: Exponenciální kvantilový graf – GP model*



Jak můžeme vyzorovat z grafu, odhadnutý model poměrně dobře koresponduje s reálnými daty, neboť hodnoty seřazené proměnné  $R_i$  se pohybují těsně kolem jednotkové přímky, vyjma několika posledních koncových bodů. Tímto se graficky potvrdilo, že volba prahové proměnné, jakožto funkce ceny vozidla, se osvědčila, a takto vyjádřenou závislost lze tedy doporučit pro modelování extrémních individuálních pojistných škod v rámci havarijního pojištění osobních vozidel.

#### 4.4 POROVNÁNÍ REGRESNÍCH MODELŮ

Dle již dříve popsaných teoretických východisek obou přístupů k modelování pojistných škod můžeme předpokládat, že lépe by reálným datům měl odpovídat smíšený model  $M_2$ , neboť umožňuje zohlednit extrémní ztráty. Jak je tomu v případě analyzovaného vzorku dat, zjistíme v následujících odstavcích, ve kterých se snažíme posoudit kvalitu modelu různými metodami. Jelikož je oblast srovnání regresního modelu  $M_1$  a regresního modelu  $M_2$  poměrně obtížná, je nezbytné provést určité zjednodušující kroky při vyhodnocování jejich přesnosti.

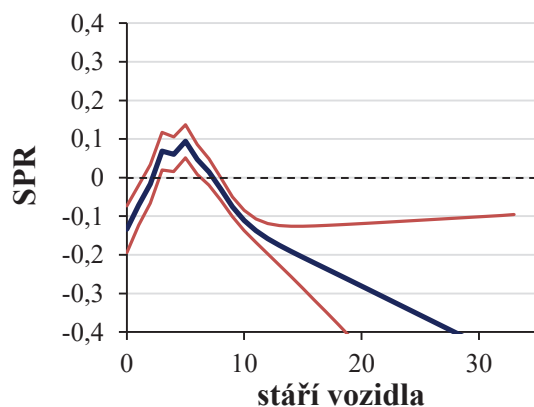
V rámci srovnání se tedy nejdříve zaměříme na grafické porovnání Pearsonových reziduí odpovídajících modelu  $M_1$  a log-gamma části smíšeného modelu  $M_2$ . Nemůžeme zde pro zhodnocení zahrnout Pearsonova rezidua celkového smíšeného regresního modelu, neboť pro část zobecněného Paretova modelu jsou rezidua počítána jiným způsobem, a tudíž by nebylo matematicky správné se snažit sloučit je do hodnot celkových reziduí. V další fázi posouzení kvality modelů jsou analyzována hrubá rezidua, která lze již získat i pro celkový smíšený model. Jako doplňkové kritérium je pak zvolen výpočet směrodatné odchylky očekávaných hodnot pojistných škod na základě odhadnutých modelů.

#### 4.4.1 ANALÝZA PEARSONOVÝCH REZIDUÍ

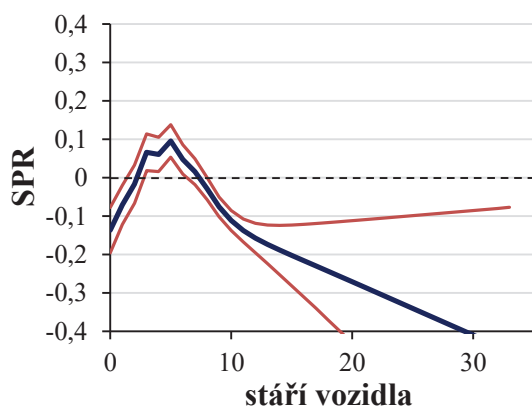
Jak již bylo výše zmíněno, Pearsonova rezidua počítáme pro celý regresní model  $M_1$  a pro log-gamma část smíšeného modelu  $M_2$ , a to dle vztahu (3.56). Takto získaná rezidua je nezbytné pro lepší grafické posouzení vyhladit, což se provádí pomocí funkcí integrovaných v programu STATA. Pearsonova rezidua jsou vyhlazována vůči rizikovým faktorům, které mají charakter spojitých veličin, tudíž vůči faktorům stáří vozidla (*agecar*), věk majitele vozidla (*age*man), cena vozidla (*carprice*), síla motoru (*kwvol*) a počet obyvatel města (*nocit*).

Jelikož odhadnuté parametry obou log-gamma modelů (modelu  $M_1$  a log-gamma části modelu  $M_2$ ) se od sebe příliš neliší, stejně jako vektor vysvětlovaných proměnných, neboť extrémní ztráty se vyskytují pouze v případě 400 smluv, lze očekávat, že vyhlazená Pearsonova rezidua (*SPR*) budou mít velmi podobný průběh jak v případě modelu  $M_1$ , tak v případě části modelu  $M_2$ . Pro každý model jsou tedy sestrojeny grafy vyhlazených reziduí vůči jednotlivým rizikovým faktorům a jsou postupně zachyceny v Graf 4.3 – Graf 4.14 níže. Modrou linií jsou zaznamenána příslušná vyhlazená Pearsonova rezidua, červené linie poté značí dolní a horní hranici konfidenčních intervalů za předpokladu 95% hladiny spolehlivosti.

Graf 4.3: *SPR* modelu  $M_1$  – *agecar*

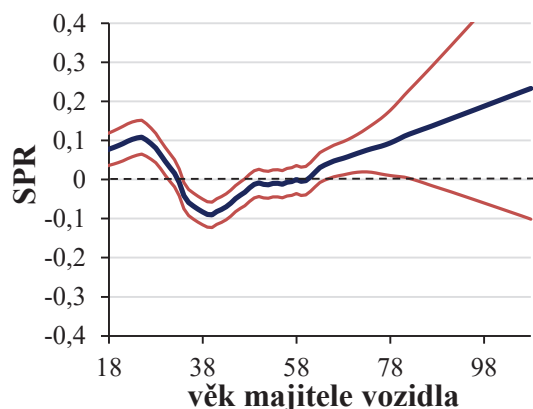


Graf 4.4: *SPR* části modelu  $M_2$  - *agecar*

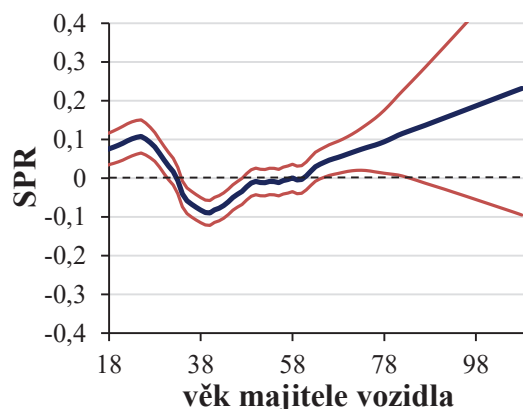


Z těchto dvou grafů lze vypožorovat, že vývoj vyhlazených reziduí je v obou případech téměř totožný, přičemž lze očekávat, že tomu tak bude i pro další spojité rizikové faktory. Podíváme-li se blíže na průběh reziduí, zjistíme, že se pohybují v rozmezí pásma od -0,1 do 0,1 přibližně v 95 % všech případů, za hraniční hodnotu lze považovat stáří vozidla na úrovni 10 let, za tímto bodem se již rezidua výrazněji odchylují od nulové linie. Tato mez byla překročena zhruba u 600 zkoumaných pojistných situací.

Graf 4.5: SPR modelu  $M_1$  – ageman

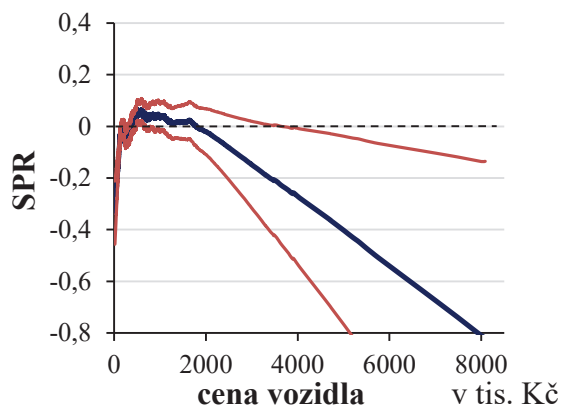


Graf 4.6: SPR části modelu  $M_2$  – ageman

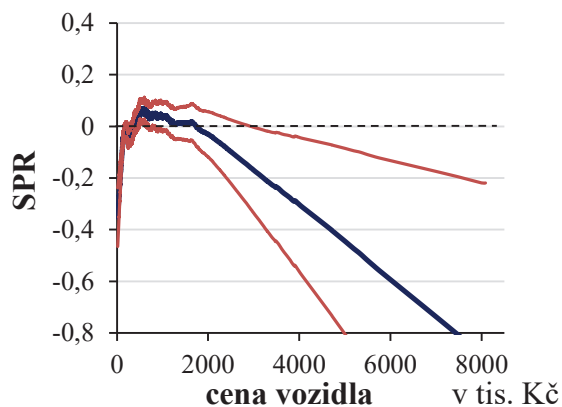


Na základě výše uvedených grafů lze opět potvrdit vysokou podobnost mezi zkoumanými modely, kdy nyní již ani pouhým okem nerozeznáme rozdíl mezi průběhem vyhlazených reziduí. Pokud se zaměříme na kvalitu modelů, pak můžeme tvrdit, že i v případě faktoru *ageman* se rezidua vyvíjí poměrně dobře, kdy u 95 % všech analyzovaných pojistných smluv rezidua nepřekračují pásmo  $\pm 0,1$ . Konkrétně 95,26 % všech zkoumaných hodnot faktoru věku majitele vozidla leží v rozmezí od 24 do 74 let včetně.

Graf 4.7: SPR modelu  $M_1$  – carprice

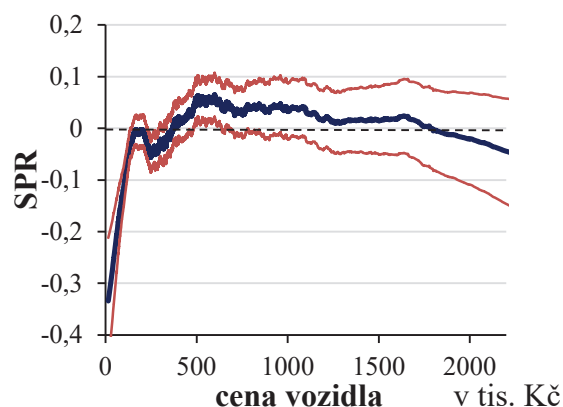


Graf 4.8: SPR části modelu  $M_2$  – carprice

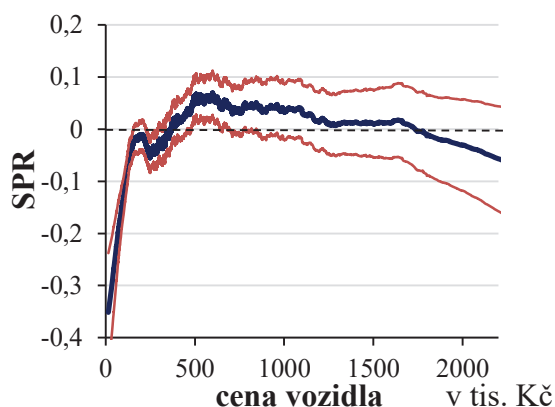


Opět i v případě zachycení faktoru ceny vozidla se vývoj reziduí u obou analyzovaných modelů velmi podobá. Na první pohled se může jevit, že vyhlazená rezidua se již výrazně odchyľují od nulové linie, nicméně vyšší cenu vozidla přesahující úroveň 2 000 000 Kč lze shledat pouze u 227 pojistných případů z celkových 22 274 pozorovaných hodnot. Z tohoto důvodu je vhodné změnit měřítko grafů a přiblížit více počáteční rezidua, viz Graf 4.9 a Graf 4.10.

Graf 4.9: Detail SPR modelu  $M_1$  – carprice

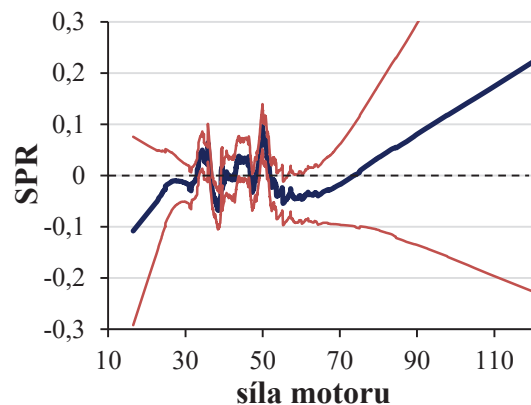


Graf 4.10: Detail SPR části  $M_2$  – carprice

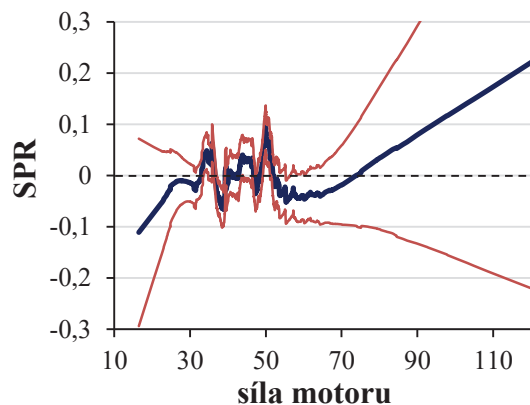


Z těchto dvou grafů je již zřejmé, že oba modely lze opět považovat za přesné a kvalitní, neboť vyhlazená Pearsonova rezidua v převážné většině náleží do intervalu od -0,1 do 0,1. Pouze pokud se ceny vozidel pohybují na úrovni 112 000 Kč a níž, pak se rezidua výrazněji odchyľují od nulové linie, nicméně to nastává pouze u 2,79 % všech individuálních pojistných smluv. Z těchto závěrů tedy vyplývá, že i v případě vyhlazení reziduí vůči rizikovému faktoru carprice více než 95 % všech reziduí spadá do pásma  $\pm 0,1$ , čímž se potvrzuje kvalita a přesnost obou uvažovaných modelů.

Graf 4.11: SPR modelu  $M_1$  – kwvol



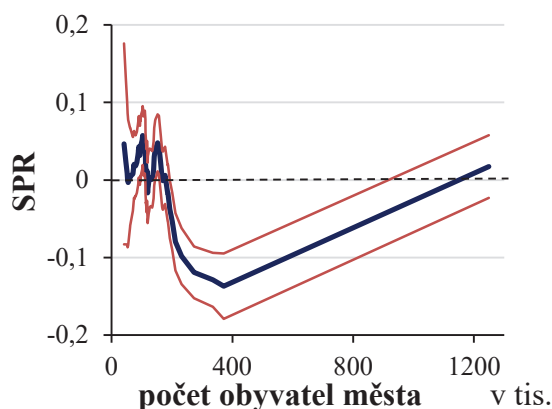
Graf 4.12: SPR části modelu  $M_2$  – kwvol



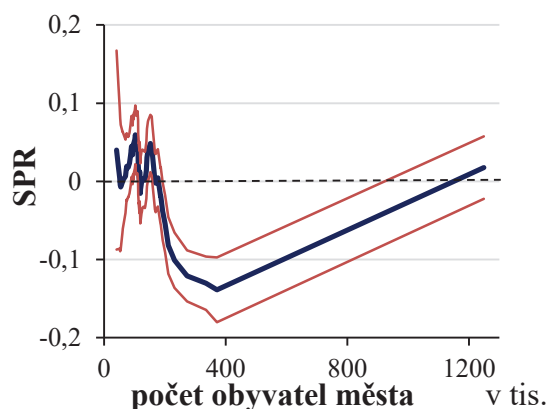


Hodnotíme-li rezidua vyhlazená vůči faktoru síla motoru, pak u obou modelů lze opět jen velmi těžko rozeznat rozdíl v jejich vývoji. Pearsonova rezidua leží blízko v okolí nulové linie, více se odchyľují až za hranicí *kwvol* na úrovni 90. Větší silou motoru však disponovalo jen 19 vozidel v rámci celkového uvažovaného počtu pojistných případů. Mimo pásmo  $\pm 0,1$  se pak pohybují hodnoty pouhých 17 vypočtených reziduů. Na základě těchto skutečností můžeme říci, že model  $M_1$  a log-gamma část modelu  $M_2$  dobře odpovídají zkoumaným pojistným datům a jsou vhodné pro modelování individuálních pojistných škod.

Graf 4.13: SPR modelu  $M_1$  – nocit



Graf 4.14: SPR části modelu  $M_2$  – nocit



Jak vidíme z obou grafů znázorněných výše, přibližně do 200 000tisícové úrovně počtu obyvatel rezidua kolísají velmi blízko nulové hodnotě, avšak poté klesne hodnota vyhlazených Pearsonových reziduů až k bodu -0,14, ze kterého následně jednorázově přeskočí zpět na úroveň blízkou 0. Samotná kategorie počtu obyvatel města je zajímavá, jelikož v České republice je mezi největším městem Prahou a druhým největším městem Brnem rozdíl přesahující 800 000 obyvatel, což se také projevilo výrazným skokem mezi dvěma body v grafech výše, kdy mezi nimi nelze spatřovat žádná kolísání a vývojové tendence reziduů. I tyto dva grafy pouze potvrzují přesnost log-gamma modelu  $M_1$  a log-gamma části smíšeného modelu  $M_2$ , neboť vyhlazená rezidua se pohybují velmi blízko nulové linie ve všech analyzovaných případech.

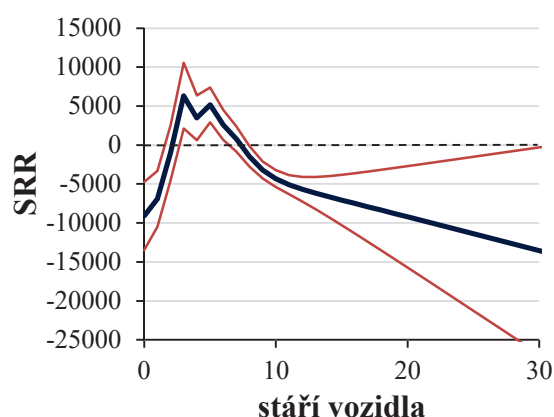
#### 4.4.2 ANALÝZA HRUBÝCH REZIDUÍ

Druhou možností zhodnocení přesnosti modelu, která je v této práci využita, je porovnání vyhlazených hrubých reziduů (*SRR*) zjištěných jak pro log-gamma model  $M_1$ , tak pro smíšený model  $M_2$ . Hrubá rezidua získáme dle vztahu (3.55), přičemž v případě smíšeného modelu střední hodnotu ztráty stanovíme dle vzorce (3.53). Velikost pravděpodobnosti  $p$  byla určena empiricky jako podíl extrémních ztrát vůči celkovému počtu

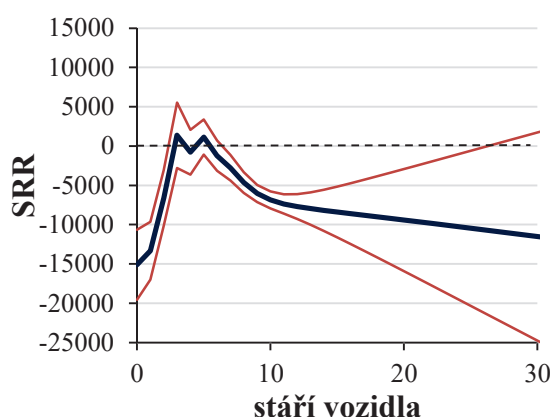
individuálních pojistných škod na úrovni 0,0180. Vypočtená hrubá rezidua opět vyhladíme pomocí programu STATA, a to vůči stejným spojitým rizikovým faktorům jako u analýzy Pearsonových reziduí.

Na základě průběhu vyhlazených hrubých reziduí následně určíme, který z posuzovaných modelů je přesnější a lépe odpovídá zkoumaným datům. Vytvořené grafy vyhlazených hrubých reziduí jsou postupně zaznamenány v Graf 4.15 – Graf 4.26 níže, přičemž vývoj vyhlazených reziduí sledujeme opět pomocí modré linie, červené linie poté představují dolní a horní hranici příslušných konfidenčních intervalů.

Graf 4.15: SRR modelu  $M_1$  – agecar

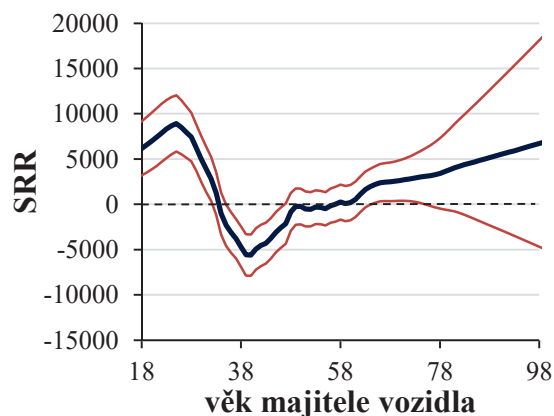


Graf 4.16: SRR modelu  $M_2$  – agecar

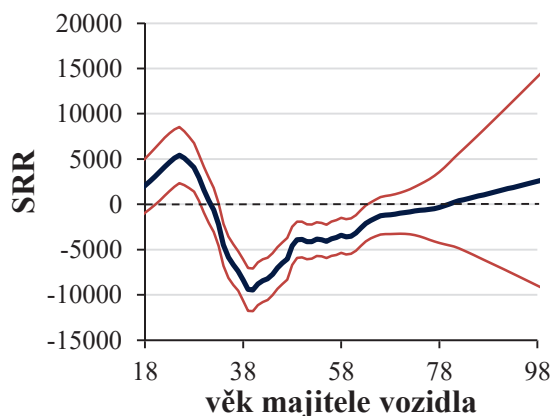


Pokud se zaměříme na srovnání vyhlazených hrubých reziduí pro faktor *agecar*, pak si na první pohled všimneme, že *SRR* u modelu  $M_2$  jsou níže položená, což svědčí o přecenění skutečných ztrát tímto modelem. Hrubá rezidua se stejně jako Pearsonova rezidua začínají výrazněji odchylovat od nulové linie přibližně za hranicí 10 let stáří vozidla, avšak v případě smíšeného modelu je toto odchylení více pozvolné, což dokládá lepší zohlednění extrémních ztrát tímto modelem.

Graf 4.17: SRR modelu  $M_1$  – ageman

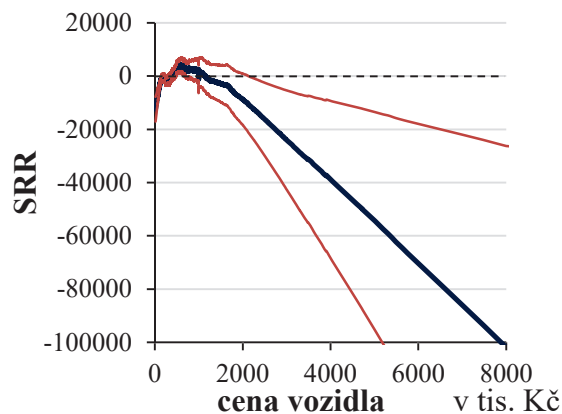


Graf 4.18: SRR modelu  $M_2$  – ageman

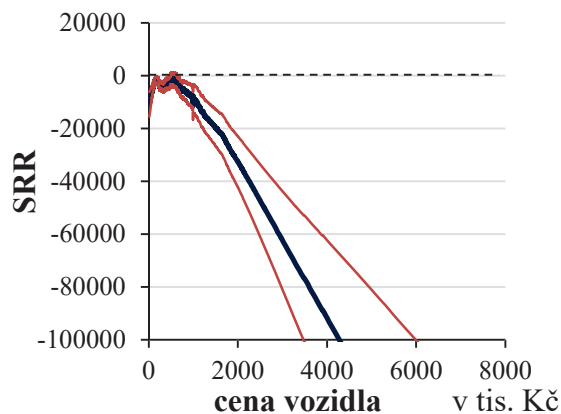


I v případě vyhlazení hrubých reziduí vůči proměnné *ageman* lze zaznamenat, že *SRR* modelu  $M_2$  leží opět na nižší úrovni než u modelu  $M_1$ , nicméně můžeme tvrdit, že v případě krajních hodnot faktoru věk majitele vozidla smíšený regresní model lépe odhaduje skutečné ztráty. Navzdory tomuto faktu se ale regresní log-gamma model ve většině pojistných případů jeví jako přesnější varianta, což je potvrzeno i součtem odchylek křivky vyhlazených hrubých reziduí od nulové linie.

Graf 4.19: *SRR* modelu  $M_1$  – *carprice*

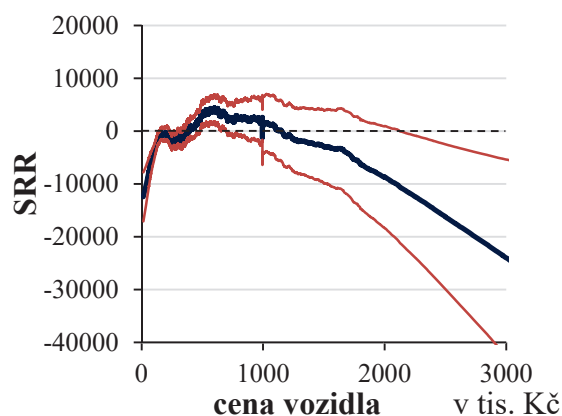


Graf 4.20: *SRR* modelu  $M_2$  – *carprice*

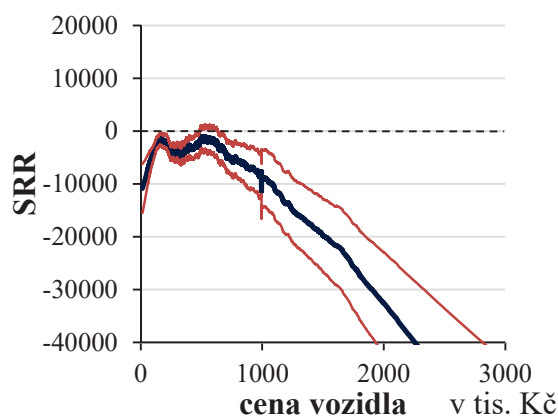


Z výše zobrazených grafů je zřejmé, že vyhlazená hrubá rezidua se v rámci faktoru *carprice* vyvíjejí lépe v případě modelu  $M_1$ , kdy se přibližně do hranice 2 mil. Kč pohybují poměrně blízko kolem 0, což je doloženo i detailním zachycením těchto grafů v Graf 4.21 a Graf 4.22 níže.

Graf 4.21: Detail *SRR* modelu  $M_1$  – *carprice*



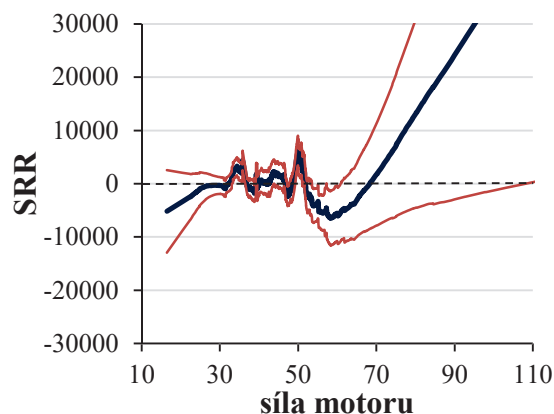
Graf 4.22: Detail *SRR* modelu  $M_2$  – *carprice*



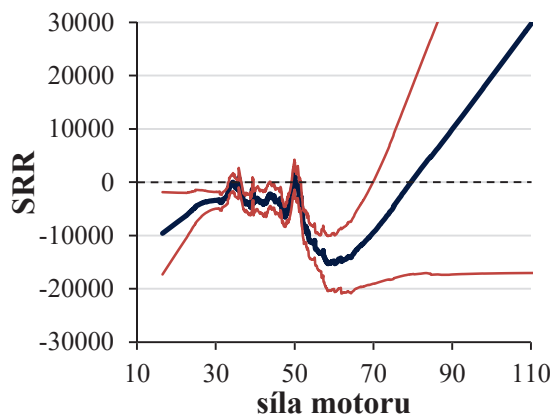
Jak lze tedy vypořádat z detailního zachycení úvodních částí křivek obou grafů, v případě modelu  $M_1$  vypočtená rezidua kolísají kolem nulové hranice, kdežto vyhlazená hrubá rezidua modelu  $M_2$  jsou pro všechny hodnoty ceny vozidla záporná, navíc se více a

rychleji odcházejí od nulové linie. Tuto skutečnost lze odůvodnit vyššími očekávanými ztrátami vypočtenými dle modelu  $M_2$  než jsou ztráty skutečné.

Graf 4.23: SRR modelu  $M_1$  – *kwvol*

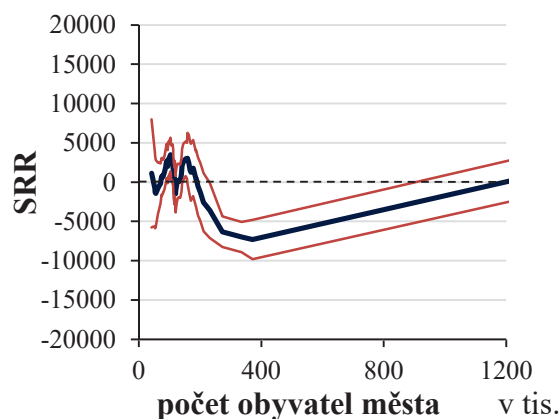


Graf 4.24: SRR modelu  $M_2$  – *kwvol*

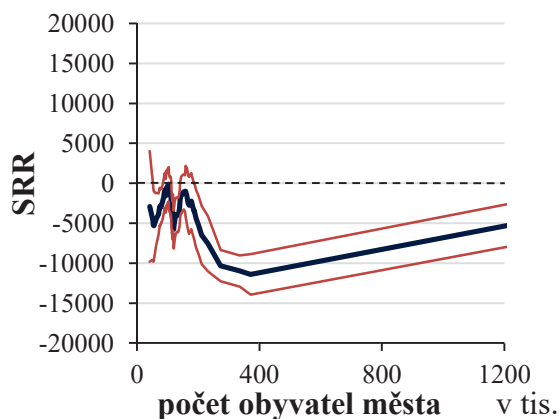


Pokud se zaměříme na vyhlazená rezidua v rámci faktoru *kwvol*, tak lze jednoznačně říci, že regresní model  $M_1$  lépe odpovídá pozorovaným datům, jelikož SRR kolísají kolem nulové hodnoty a více se odchylojí až při vysokých hodnotách vyjadřujících sílu motoru, kterými ale vozidla disponují pouze v případě několika málo zkoumaných pojistných smluv. Vyhlazená hrubá rezidua smíšeného modelu  $M_2$  jsou opět celkově níže položená, ale jejich vývoj je podobný jako u reziduí prvního modelu.

Graf 4.25: SRR modelu  $M_1$  – *nocit*



Graf 4.26: SRR modelu  $M_2$  – *nocit*



Analyzujeme-li vývoj SRR v rámci posledního zkoumaného rizikového faktoru *nocit*, zjistíme, že i zde lze usuzovat o lepší přesnosti a vhodnosti modelu  $M_1$  než modelu  $M_2$ . Křivka vyhlazených reziduí smíšeného modelu má podobný průběh jako křivka reziduí log-gamma modelu, nicméně je opět níže položená a více se tudíž odchyloje od nulové linie.

#### 4.4.3 ANALÝZA SMĚRODATNÉ ODCHYLKY

Poslední možností srovnání modelů, která byla zvolena pouze jako doplňkové kritérium, je výpočet směrodatné odchylky očekávaných individuálních pojistných škod odhadnutých na základě regresních modelů  $M_1$  a  $M_2$ . Pomocí funkce dostupné v programu MS Excel byla směrodatná odchylka očekávaných pojistných škod modelu  $M_1$  stanovena ve výši 22 653,47 Kč a modelu  $M_2$  na úrovni 26 769,43 Kč. Hodnotou tohoto kritéria je tedy rovněž potvrzeno, že regresní log-gamma model  $M_1$  je vhodnější použít na námi zkoumaná pojistná data než regresní smíšený model  $M_2$ .

#### 4.5 SOUHRNNÉ ZHODNOCENÍ REGRESNÍCH MODELŮ

Na základě teoretických východisek obou přístupů k modelování individuálních pojistných škod se očekávalo, že kvalitnějším a přesnějším modelem bude smíšený regresní model lépe zohledňující výskyt extrémních pojistných škod. Dle závěrů předešlé kapitoly to však tvrdit nelze, neboť regresní log-gamma model  $M_1$  se celkově projevil jako vhodnější pro modelování zkoumaných dat, a to i přes to, že smíšeným modelem byly lépe zohledněny extrémní pojistné škody. V následujícím textu podrobněji rozebereme možné příčiny těchto skutečností.

Pokud se blíže zaměříme na jednotlivé části smíšeného modelu, zjistíme, že samostatně je lze označit jako poměrně přesné a kvalitní modely, které dobře odhadují daná pojistná data. Zobecněný Paretův model, kterým bylo modelováno 400 extrémních individuálních pojistných škod, lze shledat jako přesný, což je graficky doloženo v Graf 4.2 výše. Rovněž z výsledků analýzy Pearsonových reziduí lze usuzovat, že i log-gamma část smíšeného modelu je nastavena správně, neboť vyhlazená rezidua kolísají kolem nulové linie a výrazněji se odchylují pouze v případě extrémních hodnot rizikových faktorů, které ve všech hodnocených případech tvoří maximálně 5 % z celkového datového vzorku čítajícího 22 274 pozorování. Pokud bereme v úvahu tyto skutečnosti, pak nedokonalosti smíšeného modelu nelze hledat v nastavení jeho jednotlivých částí ani v postupu určení prahové proměnné. Jedinou možností tedy zůstává způsob spojení těchto dvou částí v jeden celkový model, tedy zejména v určení pravděpodobnosti  $p$ . Tato pravděpodobnost byla vypočtena empiricky na konstantní úrovni 0,0180. Nejspíš právě v konstantní úrovni pravděpodobnosti  $p$  můžeme tedy shledat hlavní příčinu horších celkových výsledků modelu  $M_2$  než modelu  $M_1$ , které se projeví v rámci analýzy vyhlazených hrubých reziduí, kdy rezidua smíšeného modelu byla více vzdálená od nulové linie než rezidua log-gamma modelu. Toto tvrzení je pak podpořeno i

hodnotou doplňkového kritéria směrodatné odchylky, která pro model  $M_2$  vyšla vyšší než pro model  $M_1$ .

Zhodnotíme-li tedy přesnost pouze regresního log-gamma modelu  $M_1$ , jakožto lepšího z analyzovaných modelů, pak můžeme tvrdit, že jsou tímto modelem poměrně dobře modelována zkoumaná pojistná data, což je doloženo analýzou jak vyhlazených Pearsonových reziduí, tak vyhlazených hrubých reziduí. Vyhlazená Pearsonova rezidua se ve většině hodnocených případů dokonce pohybovala v pásmu velmi blízkém nulové hodnotě, tedy v pásmu  $\pm 0,1$ . Pouze v případě extrémně velkých či extrémně nízkých hodnot jednotlivých rizikových faktorů se rezidua vzdalovala od nulové linie. Jelikož tyto extrémní situace v rámci daných rizikových faktorů nenastaly ani u 5 % všech zkoumaných pojistných případů, můžeme tento regresní model  $M_1$  považovat za přesný a vhodný model pro modelování individuálních pojistných škod. Pokud se zaměříme na grafické zachycení vyhlazených hrubých reziduí regresního modelu  $M_1$ , zjistíme, že *SRR* se pohybují rovněž poměrně blízko nulové linie, kromě reziduí zohledňujících extrémní hodnoty jednotlivých rizikových faktorů, které ale nejsou za předpokladu 95% hladiny spolehlivosti příliš důležité. Rovněž na základě výsledků analýzy hrubých reziduí tedy můžeme regresní log-gamma model označit za kvalitní a přesný.

Závěrem této diplomové práce je nezbytné uvést, že jak regresní log-gamma model, tak regresní smíšený model lze uplatnit v pojistné analýze v rámci modelování individuálních pojistných škod, čímž je doložen již známý předpoklad, že gamma rozdělení i Paretovo rozdělení pravděpodobnosti se jeví jako vhodné pro tuto problematiku. Regresní smíšený model by avšak bylo žádoucí ještě dále rozšířit zavedením variabilní pravděpodobnosti  $p$ , což by poté mělo mít vliv na zlepšení hodnocených charakteristik modelu. Výsledky této diplomové práce by se tudíž dále mohly rozpracovat o výpočty očekávaných ztrát celkového modelu za předpokladu variabilní pravděpodobnosti  $p$ , kterou lze určit například logistickou regresí, a o následné porovnání se smíšeným modelem sestaveným na základě empiricky určené konstantní pravděpodobnosti  $p$ .

## 5 ZÁVĚR

V rámci této diplomové práce se zabýváme problematikou analýzy pojistných dat, konkrétně se zaměřujeme na modelování individuálních pojistných škod v oblasti havarijního pojištění. Cílem práce je tedy vytvořit a odhadnout regresní model pro modelování pojistných škod spojených s danou smlouvou, přičemž přesnost a kvalita daného modelu musí být ověřena a vyhodnocena.

Obsahem druhé kapitoly této práce je teoretické vymezení procesu řízení pojistných rizik, včetně nastínění terminologie či aktuální situace v oblasti risk managementu. Třetí kapitola je věnována metodice nezbytné k vytvoření regresního modelu a ohodnocení jeho přesnosti, blíže jsou zde popsány principy zobecněných lineárních modelů, teorie extrémních hodnot a reziduální analýzy.

Ve čtvrté kapitole je již zachycen postup modelování velikosti individuálních pojistných škod. Nejdříve je nutné přesně vymezit pojistná data, se kterými pracujeme. Následně jsou v této části vytvořeny dva regresní modely, konkrétně log-gamma model a smíšený model, a provedeny odhady jejich parametrů. Přesnost obou regresních modelů je na závěr posouzena a vyhodnocena zejména na základě výsledků analýzy Pearsonových a hrubých reziduí.

Prvním krokem je tedy charakteristika souboru zkoumaných pojistných dat. K dispozici máme vzorek dat zahrnující jednotlivé pojistné škody společně s individuálními charakteristikami typickými pro danou smlouvu. Celkem je analyzováno 22 274 individuálních pojistných škod, přičemž jako rizikové faktory uvažujeme stáří vozidla (*agecar*), věk majitele vozidla (*ageman*), cenu vozidla (*carprice*), sílu motoru (*kwvol*), počet obyvatel města (*nocit*), pohlaví pojistníka (*gender*), formu vlastnictví vozidla (*company*) a typ paliva (*fuel*). Poslední tři zmiňované faktory jsou vedeny jako kategorické proměnné. Na základě těchto hodnot je následně konstruován regresní model tak, aby co nejlépe a nejpresněji odpovídal skutečnosti. Pro modelování pojistných škod je zvoleno gamma rozdělení pravděpodobnosti a smíšené rozdělení pravděpodobnosti za předpokladu gamma rozdělení a zobecněného Paretova rozdělení.

V další fázi práce se již soustředíme na vytvoření regresního modelu za předpokladu gamma rozdělení pravděpodobnosti. Pomocí jednorozměrné analýzy a statistického Z-testu nejdříve ověříme statistickou významnost jednotlivých uvažovaných rizikových faktorů. Na základě výsledků jsou všechny rizikové faktory shledány jako statisticky významné, a jsou tudíž zahrnuty do modelu jako vysvětlující proměnné. Regresní model je pak konstruován za využití konceptu zobecněných lineárních modelů, kdy pomocí metody maximální



věrohodnosti odhadujeme parametry gamma rozdělení, konkrétně parametr střední hodnoty vyjádřený prostřednictvím log-link funkce a konstantní parametr disperze. Pro takto odhadnutý regresní log-gamma model v práci používáme označení symbolem  $M_1$ .

Dále je v rámci praktické části uplatněna teorie extrémních hodnot, konkrétně tedy přístup Peak over the threshold, při vytváření regresního modelu zachycujícího vztah mezi výší individuální pojistné škody a jednotlivými rizikovými faktory. Klíčovým bodem je určení prahové proměnné, která byla pro účely této práce zvolena jako funkce ceny vozidla, a její velikost se tedy mění s každou smlouvou. Na základě velikosti prahové proměnné je následně potřeba rozdělit datový soubor pojistných škod na škody nepřesahující výši dané prahové proměnné a na škody převyšující prahovou proměnnou. Tímto způsobem získáme v podstatě dva vektory, kdy jeden zachycuje „běžnou“ hodnotu pojistných škod a ve druhém jsou zahrnuty pouze extrémní ztráty. Škody nepřesahující prahovou proměnnou jsou modelovány log-gamma modelem, kdežto extrémní škody jsou modelovány zobecněným Paretovým modelem. K odhadu jednotlivých parametrů obou modelů, kdy v případě Paretova modelu odhadujeme parametr polohy, tvaru a variability, je opět využita metoda maximální věrohodnosti. Po získání odhadů parametrů je dalším krokem vytvoření tzv. regresního smíšeného modelu za předpokladu smíšeného rozdělení slučujícího gamma rozdělení a zobecněné Paretovo rozdělení pravděpodobnosti. Smíšený model, jenž se skládá v podstatě z log-gamma modelu a zobecněného Paretova modelu, je v práci zjednodušeně pojmenován jako model  $M_2$ .

Hlavním předmětem zájmu této diplomové práce je pak vyhodnocení kvality jednotlivých regresních modelů, včetně jejich vzájemného srovnání. Pro posouzení jejich přesnosti je použita zejména analýza vyhlazených hrubých a Pearsonových reziduí doplněná o výpočet směrodatné odchylky.

Vyhlazená Pearsonova rezidua sledujeme pro log-gamma model  $M_1$  a log-gamma část modelu  $M_2$ . Dle grafického zachycení se oba modely velmi podobají, což je způsobeno tím, že extrémních ztrát bylo dosaženo pouze v případě 400 pojistných smluv, a to je jen nepatrná část celého souboru dat, tudíž příslušné vektory individuálních pojistných škod se od sebe příliš neliší. Vyhlazená Pearsonova rezidua mají u obou modelů požadovaný vývoj, jelikož kolísají v pásmu velmi blízkém nulové linii a výrazněji se začínají odchylovat až v případě extrémních hodnot daných rizikových faktorů. Na základě těchto závěrů můžeme shledat oba modely jako poměrně přesné a vhodné pro pozorovaná data.

Dále se zabýváme analýzou vyhlazených hrubých reziduí, která jsme schopni získat již pro celkový smíšený model. Vyhlazená hrubá rezidua jsou opět porovnávána mezi dvěma



modely pomocí grafického zachycení. Ze sestrojených grafů pak vyplývá, že log-gamma model  $M_1$  se jeví jako vhodnější a přesnější varianta pro modelování analyzovaných pojistných škod než smíšený model  $M_2$ . Toto tvrzení je také potvrzeno výpočtem směrodatných odchylek očekávaných ztrát pro oba modely. V případě log-gamma modelu je vývoj hrubých reziduí podobný jako vývoj Pearsonových reziduí, vyhlazená hrubá rezidua se tedy pohybují téměř ve všech zkoumaných případech poblíž nulové hodnoty, a to kromě extrémních hodnot, kde lze již spatřovat významnější odchylky. Pokud se zaměříme na vyhlazená hrubá rezidua smíšeného modelu, pak u všech rizikových faktorů je jejich křivka níže položená než u modelu  $M_1$ , což značí, že modelem  $M_2$  jsou očekávané ztráty nadhodnocovány oproti těm skutečným.

Pokud souhrnně zhodnotíme oba uvažované regresní modely na základě výsledků analýzy hrubých reziduí a směrodatné odchylky očekávaných ztrát, pak dojdeme k závěru, že v rámci námi zkoumaných pojistných dat se jako lepší zdá být regresní log-gamma model  $M_1$ . Zaměříme-li se však na samostatné zhodnocení obou částí smíšeného modelu, tak zjistíme, že jak log-gamma část modelu, tak také zobecněnou Paretovu část modelu lze označit za velmi přesnou a dobře odpovídající zkoumaným datům, což bylo graficky doloženo výsledným průběhem vyhlazených Pearsonových reziduí, respektive vytvořením exponenciálního kvantilového grafu. Příčinu horších celkových výsledků v rámci analýzy vyhlazených hrubých reziduí můžeme u smíšeného modelu  $M_2$  tedy hledat zejména ve způsobu spojení jeho dvou částí v jeden celek.

Shrneme-li výše uvedené závěry, pak regresní log-gamma model lze jednoznačně použít pro modelování individuálních pojistných škod, přičemž jeho přesnost a kvalita byla ověřena reziduální analýzou. Princip smíšeného regresního modelu můžeme také aplikovat na analýzu pojistných dat, nicméně by bylo potřeba zapracovat na způsobu sloučení obou rozdělení pravděpodobnosti, a tedy i modelů tak, aby lépe zachycoval skutečná data. Výsledky této diplomové práce byl také potvrzen předpoklad, že gamma rozdělení pravděpodobnosti a zobecněné Paretovo rozdělení pravděpodobnosti jsou vhodnými rozděleními pro modelování pojistných dat, hlavně tedy pojistných škod.

## SEZNAM POUŽITÉ LITERATURY

### ♦ knižní publikace

1. BEIRLANT, J., GOEGEBEUR, Y., J. SEGERS and J. TEUGELS. *Statistics of Extremes: Theory and Applications*. 1st ed. Chichester: Wiley, 2004. 490 s. ISBN 978-0-471-97647-9.
2. CIPRA, Tomáš. *Kapitálová přiměřenost ve financích a solventnost v pojišťovnictví*. 1. vyd. Praha: Ekopress, 2002. 271 s. ISBN 80-86119-54-8.
3. COLES, Stuart. *An Introduction to Statistical Modeling of Extreme Values*. 1st ed. London: Springer, 2001. 208 s. ISBN 978-1-84996-874-4.
4. DUCHÁČKOVÁ, Eva. *Principy pojištění a pojišťovnictví*. 3. vyd. Praha: Ekopress, 2009. 224 s. ISBN 978-80-86929-51-4.
5. DUCHÁČKOVÁ, Eva a Jaroslav DAŇHEL. *Teorie pojistných trhů*. 1. vyd. Praha: Professional Publishing, 2010. 216 s. ISBN 978-80-7431-015-7.
6. HARDIN, James W. and Joseph M. HILBE. *Generalized Linear Models and Extensions*. 2nd ed. College Station: Stata Press, 2007. 387 s. ISBN 978-1-59718-014-6.
7. JONG, Piet de and Gillian Z. HELLER. *Generalized Linear Models for Insurance Data*. Cambridge: Cambridge University Press, 2008. 196 s. ISBN 978-0-521-87914-9.
8. OHLSSON, Esbjörn and Björn JOHANSSON. *Non-Life Insurance Pricing with Generalized Linear Models*. 1st ed. Berlin: Springer, 2010. 174 s. ISBN 978-3-642-10790-0.
9. ŘEZÁČ, František. *Řízení rizik v pojišťovnictví*. 1. vyd. Brno: Masarykova univerzita, 2011. 222 s. ISBN 978-80-210-5637-4.

### ♦ internetové zdroje

10. ČESKÁ NÁRODNÍ BANKA. *Regulace a dohled nad pojišťovnami a zajišťovnami* [online]. ČNB [20. 2. 2016]. Dostupné z:  
[https://www.cnb.cz/cs/dohled\\_financni\\_trh/vykon\\_dohledu/postaveni\\_dohledu/pojistovny\\_zajistovny/index.html](https://www.cnb.cz/cs/dohled_financni_trh/vykon_dohledu/postaveni_dohledu/pojistovny_zajistovny/index.html)
11. KPMG. *Insurance risk and capital transformation* [online]. KPMG [22. 2. 2016]. Dostupné z:

<https://www.kpmg.com/PE/es/IssuesAndInsights/ArticlesPublications/Documents/Insurance-risk-and-capital-transformation.pdf>

12. MATHWORKS. *Modelling Data with the Generalized Extreme Value Distribution* [online]. MATHWORKS [10. 3. 2016]. Dostupné z:  
<http://www.mathworks.com/help/stats/examples/modelling-data-with-the-generalized-extreme-value-distribution.html>
13. MINISTERSTVO FINANCÍ ČESKÉ REPUBLIKY. *Legislativa: Materiály na jednání vlády* [online]. MFČR [20. 2. 2016]. Dostupné z:  
<http://www.mfcr.cz/cs/legislativa/materialy-na-jednani-vlady/1-ctvrtleti-16/materialy-na-jednani-vlady-dne-3-unora-2-23868>
14. STACK OVERFLOW. *Questions* [online]. STACKOVERFLOW [22. 3. 2016]. Dostupné z:  
<http://stackoverflow.com/questions/23196088/how-can-i-generate-an-exponential-q-q-plot-in-r>

## SEZNAM ZKRATEK

ART	Alternative Risk Transfer
ČNB	Česká národní banka
EML	Estimated Maximum Loss
EVT	Extreme Value Theory
ERM	Enterprise Risk Management
GEV	Generalized Extreme Value
GLM	Generalized Linear Models
Kč	korun českých
LC	Large Claims
LCI	Lower Confidence Interval
MAS	Maximum Amount Subject
mil.	million
MLM	Maximum Likelihood Method
NLE	Normal Loss Expectancy
PML	Probable Maximum Loss
POT	Peak over the Threshold
SC	Small Claims
SE	Standard Error
SPR	Smoothed Pearson Residuals
SRR	Smoothed Raw Residuals
UCI	Upper Confidence Interval

## Prohlášení o využití výsledků diplomové práce

Prohlašuji, že

- jsem byla seznámena s tím, že na mou diplomovou práci se plně vztahuje zákon č. 121/2000 Sb. – autorský zákon, zejména § 35 – užití díla v rámci občanských a náboženských obřadů, v rámci školních představení a užití díla školního a § 60 – školní dílo;
- beru na vědomí, že Vysoká škola báňská – Technická univerzita Ostrava (dále jen VŠB-TUO) má právo nevýdělečně, ke své vnitřní potřebě, diplomovou práci užít (§ 35 odst. 3);
- souhlasím s tím, že diplomová práce bude v elektronické podobě archivována v Ústřední knihovně VŠB-TUO a jeden výtisk bude uložen u vedoucího diplomové práce. Souhlasím s tím, že bibliografické údaje o diplomové práci budou zveřejněny v informačním systému VŠB-TUO;
- bylo sjednáno, že s VŠB-TUO, v případě zájmu z její strany, uzavřu licenční smlouvu s oprávněním užít dílo v rozsahu § 12 odst. 4 autorského zákona;
- bylo sjednáno, že užít své dílo, diplomovou práci, nebo poskytnout licenci k jejímu využití mohu jen se souhlasem VŠB-TUO, která je oprávněna v takovém případě ode mne požadovat přiměřený příspěvek na úhradu nákladů, které byly VŠB-TUO na vytvoření díla vynaloženy (až do jejich skutečné výše).

V Ostravě dne 22. dubna 2016



.....  
Bc. Šárka Marčíková

## **SEZNAM PŘÍLOH**

*Příloha č. 1: Výsledky jednorozměrné analýzy dat*

*Příloha č. 2: Statistická verifikace pomocí Z-testu*